R842-09 ERIC REPORT RESUME

ED 010 186 1-31-67 24 (REV)
RELATIONSHIPS AMONG LEARNING, PRACTICE, AND RECALL, ~~FINAL REPORT.~~ OK
BETTS, MARY L.B.
FPI18901 HARVARD UNIV., LAB. FOR RES. IN INSTRUCT., CAMBRIDGE, MASS
CRP-S-169
FPI18897 HARVARD UNIV., GRAD. SCH. OF ED., CAMBRIDGE, MASS.
BR-5-8195
- -66
EDRS PRICE MF-$0.36 HC-$8.64 216P.

*LEARNING, COLLEGE STUDENTS, *LEARNING PROCESSES, LEARNING THEORIES, *LEARNING EXPERIENCES, ABILITY, *INDIVIDUAL DIFFERENCES, CAMBRIDGE, MASSACHUSETTS

THIS STUDY CONSISTED OF A SERIES OF EXPERIMENTS DESIGNED TO YIELD DATA ON THE PARAMETERS OF LEARNING IN THESE SITUATIONS--(1) THE QUANTITY OF MATERIAL TO BE LEARNED WAS GREATER THAN USUAL, AND (2) THE SCHEDULING OF LEARNING, REVIEW TRIALS, AND THE AMOUNTS OF MATERIAL PRESENTED WERE VARIED WITHIN A DESIGN JUDGED TO APPROACH OPTIMAL SCHEDULING IN CERTAIN RESPECTS. OF PARTICULAR INTEREST WERE QUESTIONS HAVING TO DO WITH THE POSSIBLE INTERFERENCE OF ONE LIST OF PAIRED-ASSOCIATE MATERIAL WITH ANOTHER, AND THE ROLE OF INDIVIDUAL DIFFERENCES IN LEARNING ABILITY. THE EXPERIMENTS INVOLVED 9 GROUPS OF COLLEGE STUDENTS, 10 SUBJECTS IN EACH GROUP. A NUMBER OF QUESTIONS CONCERNING THE LEARNING CURVE DATA WERE ANSWERED BY USING A TREND ANALYSIS WITH COVARIANCE ADJUSTMENTS FOR ABILITY SCORES. IN ADDITION, A MULTIPLE REGRESSION EQUATION WAS USED TO ASSESS THE USEFULNESS OF SIX VARIABLES IN PREDICTING LEARNING TIMES. THE AUTHOR POINTS OUT THAT THE ANALYSES PRESENTED DO NOT DEPICT A COMPLETE PICTURE OF LEARNING. HOWEVER, MANY OF THE FINDINGS ARE NEW AND OF THEORETICAL AND PRACTICAL INTEREST. (JC)

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

FINAL REPORT

RELATIONSHIPS AMONG LEARNING, PRACTICE, AND RECALL

Cooperative Research Project No. S-169

Mary Long Burke Betts

Laboratory for Research in Instruction
Graduate School of Education
Harvard University
Cambridge, Massachusetts

1966

ED010186

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U.S. Department of Health, Education, and Welfare.

Acknowledgements

This thesis represents not only the obvious efforts of its author, but also the quiet, unstinting help and advice of numerous others. A very heartfelt "thank you" is due to Dr. Paul Kjeldergaard, Dr. Douglas Kenny, Professor Gerald Lesser, Dr. Allan Ellis, Dr. Kenneth Jones, Dr. Albert Bregman, Dr. William Young, and Professor Phillip Rulon for their help and encouragement at emotionally and statistically crucial moments. Professor A. P. Dempster generously gave a great deal of his valuable time to help with statistical problems, and his graciousness is greatly appreciated.

Miss Sally Chapin, Mrs. Livia Balling, and Mrs. Leonard Faltz carried out the experimentation cheerfully and carefully. Miss Frances Ricker provided countless secretarial assistances. Miss Marjorie Morse produced the graphs and tables as only she could have, and it is not the first time that she has come to the aid of the author superbly.

The thesis experiment was made possible by a grant from the United States Office of Education, and the support of this Office is very much appreciated.

There is one more "thank you" to be said. The phrase most inadequately expresses the author's gratitude to Professor John B. Carroll for five years of meticulous and fascinating training, for innumerable generosities and unlimited patience. A more valuable apprenticeship could not be imagined. What is good here belongs to Dr. Carroll, and to the others named above; the mistakes are all mine.

M. L. B.

TABLE OF CONTENTS

Page

ABSTRACT v

CHAPTER I: INTRODUCTION AND BACKGROUND 1

- General Introduction 1
- The Whole and Parts Method of Learning 8
- Interference Theory 15
- Ability 36
- Meaningfulness 38
- Warm-up 40
- Surroundings 41
- Summary of Research Findings 42

CHAPTER II: EXPERIMENTAL DESIGN 46

- General Problems 46
- Experimental Design - List Length 49
  - Experiment I 51
  - Experiment II 52
  - Experiment III 53
  - Experiment IV 54
- Other Variables 55
  - Ability 55
  - Recall 56
  - Errors 57
- Experimental Details 58
  - Subjects 58
  - Materials 59
  - Method 61
  - Procedure 62
  - Criterion 63

CHAPTER III: RESULTS AND DISCUSSION 65

- Introduction 65
- List Length 65
  - Experiment I 71
  - Experiment II 74
  - Experiment III 89
  - Experiment IV 98
- Ability 112
- Errors and Omissions 118
- Recall and Relearning 119
- Conclusions 125

APPENDIX A: MATERIALS USED IN EXPERIMENTAL AND PRACTICE LISTS 137

APPENDIX B: GRAPHS OF LEARNING TIMES FOR THE THREE CONTROL GROUPS AND FOR EACH OF THE SIX EXPERIMENTAL GROUPS 141

TABLE OF CONTENTS

(continued)

Page

APPENDIX C: GRAPHS OF LOG_e LEARNING TIMES FOR THE THREE CONTROL GROUPS AND FOR EACH OF THE SIX EXPERIMENTAL GROUPS 149

APPENDIX D: GRAPHS OF TOTAL LEARNING TIMES PER DAY FOR THE SIX EXPERIMENTAL GROUPS 157

APPENDIX E: GRAPHS OF THE NUMBER OF CORRECT RESPONSES AT RECALL OVER FOUR DAYS FOR ALL GROUPS 159

APPENDIX F: GRAPHS OF PERCENTAGE OF CORRECT RESPONSES AT RECALL OVER FOUR DAYS FOR ALL GROUPS 167

APPENDIX G: GRAPHS OF THE NUMBER OF CORRECT RESPONSES TO THE PRACTICE LIST OVER FIVE DAYS FOR ALL GROUPS 175

APPENDIX H: CORRELATIONS AMONG ALL LOG_e LEARNING AND RELEARNING TIMES: ALL GROUPS 183

APPENDIX I: PART V: MODERN LANGUAGE APTITUDE TEST (CARROLL AND SAPON) ADAPTATION BY DAVID HORTON 190

REFERENCES 196

VITA 203

ABSTRACT

The background problem with which this thesis was concerned is that of what may be called the "optimization of learning." In simple terms, this means that if the learner has a certain amount of material that he must learn within a given time in such a way that all parts of the lesson are to be equally well recalled at some later period, what is the optimal or most efficient way of allocating time to original learning and to practice of each part of the lesson? Classically, this has been called the "whole-part" problem in learning. Although a considerable amount of work was done on it, mainly in the earlier years of this century, the problem has never been satisfactorily solved. Now that the research in paired-associate verbal learning of the last few years has yielded much information concerning what variables are important in such learning, the time seemed to be ripe to attack the problem anew. The study contained a series of experiments designed to yield data on the parameters of learning in situations in which (a) the quantity of material to be learned was somewhat greater than has usually been the case, and (b) the scheduling of learning and review trials and the amounts of material presented or practiced in each trial were systematically varied within a design that was thought to approach optimal scheduling in certain important respects. Of considerable interest were questions having to do with the possible interference of one list of paired-associate material upon another, and with the role of individual differences in learning ability.

In brief, the experiment involved nine groups of subjects, ten subjects in each group. Subjects were paid volunteers from the Radcliffe freshman class, naive in experiments of this type. Consider first the

three control groups, each of which had to learn, to a criterion of one perfect trial, a list of a given length (16, 26, or 42 pairs in which the subject must learn the response associated with a given stimulus) on the first day of the experiment. This same list was then relearned on each of the next three successive days, with a recall and relearning session 72 hours after the fourth day's session. Times for learning and relearning were obtained in all cases, together with other data. In the experimental groups, however, subjects had not only to learn and relearn the original lists on each successive day, but also, beginning with the second day and continuing up to the fourth day, learn and relearn new lists, with, again, a learning and relearning session on the 7th day. The lengths of the interpolated lists were systematically varied in order to represent different types of learning scheduling, i.e. distribution of effort over the total amount to be learned. The experiment featured two pre-test measures of paired-associate learning ability (Part V of the MLAT and the number of correct responses on the first day's "warm-up" list). The "warm-up" sessions were repeated at the start of each day's learning. The schema below describes the subjects' tasks in the experimental groups.

Day 1	Day 2	Day 3	Day 4	Day 5
List 1	List 2	List 3	List 4	List 4
	List 1	List 2	List 3	List 3
		List 1	List 2	List 2
			List 1	List 1

The length of the various lists was the manipulated variable and time required to learn or relearn was the criterion. There were three list lengths used: 16, 26, and 42 pairs. These list lengths were chosen because their logs (to the base e) were approximately equidistant. Learning curves for the three control groups were presumed free of "interference" due to interpolated material. Lists in the experimental groups varied in terms of first list and all other lists; the second, third, and fourth lists were tested at all points but were considered "interference" lists. The schema below describes the list lengths used in the various groups. The first number refers to the length of list 1. The subscript refers to the length of lists two, three, and four.

Controls	16_{00}	26_{00}	42_{00}
Experimental	16_{16}	26_{16}	42_{16}
		26_{26}	42_{26}
			42_{42}

Log$_e$ scores for the four learning and relearning times for any one list were transformed by means of a fourth-order orthogonal polynomial, as:

$M_1 = T_1 + T_2 + T_3 + T_4$ = a measure of the "mean" of the or line

$M_2 = -3T_1 - T_2 + T_3 + 3T_4$ = a measure of slope

$M_3 = T_1 - T_2 - T_3 + T_4$ = a measure of curvature

$M_4 = -T_1 + 3T_2 - 3T_3 + T_4$ = a measure of double curvature or S-curvature

We asked the following questions of the data, using a trend analysis with covariance adjustments for ability scores, (Part V of the MLAT and number correct in five trials on the practice list of day one), and using the "M" transformations of log$_e$ time to learn as the dependent variables:

1. Do learning curves vary for the control group lists at the

three different stages?

We found that they did differ. Their means (or M_1's), their slopes (or M_2's), and their curvature (or M_3's) differed (though, because of the method used, this last was not easily interpreted).

2. A) For the groups that received equal amounts of material on all days, do learning and relearning curves for List One differ?

The differed in means and in slopes, but not in curvature. Since they are at three different list lengths, this is not surprising.

B) Do the initial lists in each of the three groups differ from their respective control-group lists?

Learning curves for the three initial lists differ from those of their respective control lists only in the overall means for learning times. This may be a result of equal list lengths, since list differentiation is made more difficult by lists of equal length. Or it may be due to the presence of other lists without being a function of their length.

3. A. If interference lists are held constant, do learning curves differ for the initial list lengths (Groups 16-16, 26-16, and 42-16)?

The curves for these lists differ in slopes and in curvature but, oddly enough, not in means.

B. Do the curves for the initial lists differ from those of their respective control lists?

The curves do not differ in any way from those of their respective control groups, except that Group 16-16 differes from Group 16-00 in means. Evidently, for subjects learning 26 and 42 pairs, 16-pair interference lists are not a problem.

4. If initial list length is held constant, and interpolated lists are varied, do initial learning curves differ?

The learning curves for the three groups which began with a 42-pair list (42-16, 42-26, 42-42) do not differ in any way. Curves for Groups 42-26 and 42-42 differ from the 42-pair control group curves in means, but otherwise there are no differences between experimental and control lists here.

5. How is ability related to the tasks given subjects?

Ability scores did not correlate highly, in general, with relearning times. They were related to initial learning times for the various lists. Intercorrelations among learning and relearning times for individual lists were not high, but correlations among initial learnings of the four lists in each group were high. Evidently, given time to learn initially, slower students "catch up" at relearning. The possibility is worth investigating.

In addition to these analyses, a multiple regression equation was worked out, using $\log_e$ time required to learn or relearn as the criterion, in order to assess the usefulness of six variables in predicting such learning times. The two best predictors of learning time were the list length and the number of times the list had already been seen. The two ability measures contributed to the correlation, but their true role will have to wait for a more thorough analysis of ability under the conditions used here. The number of items in other lists presented and the day on which the list of interest was first presented, (our "interference" variables) also contributed to the equation, although their first-order correlations with the criterion were not significant. Their role also needs clarification.

The results of the experiment suggested several possibly fruitful lines of research, particularly work with ability over time, work with longer list lengths, and work with varying amounts of practice or relearning after the subject learns a list to criterion initially. This last was suggested by the final recall and relearning scores, which indicated that students had remembered a satisfactory amount and had, in some cases, overlearned.

It appeared that, while much more needs to be done before results can be put to any practical use, the approach to optimization of learning used in this study has value and should be worked with further.

CHAPTER I

Introduction and Background

More than one educator has remarked upon the similarity of the task involved in paired-associate learning to the tasks students perform in school. While such skills and techniques as "critical thinking" and "discovery method" are much in the foreground today, the fact remains that a great deal of the material students have to master is of a paired-associate nature--mathematical formulae and their applications, historical events and dates, correspondences between sounds and letters in reading, and the most frequently mentioned example, foreign language vocabulary. The variables that are of interest to the researcher in paired-associates, such as amount of material, subjects' abilities, amount of time needed to learn given materials, difficulty or "meaningfulness" levels of materials and their relationships to learning, and a score or more of others, are also of interest to educators. Often the same questions are raised by educators and psychologists in the laboratory. Yet, until a few years ago, results from the laboratory studies were seldom reflected in educational practices or materials. Textbooks in foreign languages, for example, were and still are organized according to editors' notions of the structure of the language rather than the structure of the learner (some of them do not appear to be organized on any discoverable basis). This apathy on the part of educators is, in part, understandable; the usual classroom situation, in which the teacher is confronted with thirty or more students and armed only with a text, is not very similar to the highly controlled laboratory situation, where, usually, the experimenter works with individual subjects and considers only a few variables at a time. The experimenter sees his subject only a few times, perhaps only once. The subject is usually motivated by a desire to earn money or a desire to pass the course for which his services are a requisite. Few of the confusions resulting from the interactions of thirty individuals over an extended period of time are present to hamper the experimenter as they may the teacher. And the experimenter may, if he wishes, gear the material to the subject's ability; he has the time and the tools to measure his subject and adjust to him. The teacher can, at best, approximate this personal tailoring of material to student. In short, while the classroom with its confusions and its possibilities may have a great deal of merit for certain kinds of teaching, it is unlike the experimental situation in so many ways that it is not surprising that findings in experimental studies have not been used extensively in educational settings.

This situation has changed markedly in the last few years. New techniques in education, such as language laboratories, programmed instruction, and a number of detailed curricula worked out according to psychological principles as well as demands of the subject areas, offer to educators a control over the learning behaviors of students that was not possible in the ordinary classroom. According to proponents of programmed instruction, for example, it will be possible to tailor lessons to the ability of the individual student. Even the textbooks can be organized much more efficiently than they have been. Science curricula, reading programs, and other aspects of learning are being reworked with information about learning in mind. These new techniques bring a greater degree of control to educators; it may now be possible to manipulate learning variables in much the same way that the experimenter does. They also bring a need for much greater specificity in our knowledge of how students learn, a need for much more information about the parameters of learning, in a given situation or lesson and over longer periods of time--weeks, months, semesters.

We have already commented on the fact that many of these new educational techniques raise questions about the same variables that experimenters in paired-associate research have studied. Paired-associate research has been guided in the past by the curiosity of the researchers and the logical structure of the body of information on which they were working. For whatever reasons, the information presently available about parameters of learning paired-associates has some large lacunae from the educator's point of view. Certainly, educators are as interested in students' abilities, relationships among difficulty levels of material, amounts of material, potential interference and the like as experimenters have been. But some of the differences between experimental studies as they have been conducted to the present and the educational setting for paired-associate learning must be considered if learning in the schools is to benefit from theory. The most striking differences between the classic paired-associate studies and the educational situation are these:

1. A limited amount of time is available for any one academic subject. Research in learning speed has not usually considered the problem of optimizing learning within a constant amount of time per day.
2. Forgetting is, theoretically at least, not allowed. Teachers do not give a lesson and then allow students to go home and forget it as subjects in laboratory studies often do. Certainly some studies have dealt with retention times, but not over long periods of time and large amounts of material.
3. The amount of material schools hope to teach is large. Few paired-associate studies have attempted to teach an amount of material that would approximate even the few hundred words considered basic to foreign language vocabulary, for example.

Certainly few, if any, studies have dealt with all three of these characteristics at once. What sorts of learning behaviours occur when students are asked to learn a large amount of material in a constant amount of time per day over weeks or months, and asked, further, to remember all of it? Educators cannot change the parameters of the situation, at least not beyond certain limits.

The material has to be learned and remembered, and only so much time is available for study.

Put more constructively, the question becomes, "How can we make maximum use of students' time?" As Carroll (1963) has put it,

> What is wanted are statements such as "If you want to teach a vocabulary of size n, you may expect the student to spend x hours in attaining mastery, studying in such-and-such a manner." (p. 1076)

In programmed instruction, for instance, research has been concentrated primarily on the perfection of individual programs and on the development of techniques for programming individual subject areas. Researchers have sought answers to such questions as "How do you program simple spelling or reading or math concepts for young children whose reading level is, as yet quite low? How do you elicit correct answers in a foreign language program?" and the like. Few studies have been devoted primarily to developing sequences of programs, such that students' time is used efficiently. There have been some sequences of programs, but their purpose was not optimization of learning within set time limits; generally the experimenters did not keep data on such variables (Carroll's study of programmed Chinese (1963) is a noteworthy exception). Even where programmers have paid some attention to the relationship between learning and recall over time, their efforts have been, of necessity, a matter of trial and error. If this new tool, and other similar techniques are to be of real use in education, we need to be able to conduct the development of programs, curricula, and the like within a framework of theory; we need to be able to predict the shapes of relationships among such variables as amount of material to be learned, amount of practice needed, and students' abilities so that our research will not be totally trial and error.

The issue, then, is how to optimize the student's learning, how to find the best balance between time spent practicing material already presented to him and time spent learning new items, such that the student will progress with all due speed through the material without forgetting what he has already been taught, and without needlessly wasting his time. Most assuredly, this statement smacks of Utopia, particularly when we consider the current state of our knowledge about the variables involved in learning in the school situation. If we are to optimize the student's learning, we need to know a great deal more about learning behaviors over time, over large amounts of material, and under the restriction that material learned must be remembered. For the purposes of this discussion, let us hypothesize a student who is learning the vocabulary of a foreign language. If we wished to create a "curriculum" for him, which would meet the above requirements as to efficiency, we would need answers to such questions as:

1. How many items should there be in his first lesson?
2. How many items should there be in his second lesson, his third, and so on? We would wish to organize these groups

of material so that the student would not waste time and so that we did not confuse or damage yesterday's lesson by presenting too much new material today.

3. How much practice does our student need on each of the lessons in order to insure the presence of each item in his repetoire at all times? At what point may we begin to drop items, assuming that they are so well-learned that the occasional practice afforded by reading the word in context will suffice to maintain them?
4. Is the student's learning rate steady? That is, will the student find it more difficult to learn and remember items after a point in time; will the cumulative effect of several past lessons cause confusion and slower learning of lists of the same length as the earlier lists?
5. How is the student's ability related to the above questions? What abilities are required for paired-associate learning, how do we test for them, and how are differing ability levels reflected in the learning behaviors of the students?
6. How are all these factors or variables related? Can we predict the amount of time a given student needs to learn a given amount of material (where learning implies remembering) on any day, at any point in the sequence of lessons? If the student is to make the best use of a constant amount of time available per day for learning, we

must be able to balance amounts practice and new learning efficiently.

If we wish to add to the body of information in verbal behavior, or more specifically, paired-associate research, it would be foolish to duplicate what has already been done. It would be helpful in planning our research if we could find any indications of what our hypothetical student might do under various conditions of learning, insofar as experimentation has dealt with his situation. And, it would be advantageous to construct our work so that it would "fit in" with, or, if possible, extend the current knowledge about paired-associates in some systematic way. For all these reasons, it would be helpful to look first at what is already known about the variables implied in the six questions above. (The discussion which follows is not necessarily intended as a critical review of the literature; it is, rather, a search for whatever might be of relevance to our particular questions. Only those studies whose findings merit serious consideration have been included, with a few exceptions--notably the research in whole-part learning).

The Whole and Parts Method of Learning

One of the oldest and earliest attempts to make learning efficient is the body of research usually labeled "the whole-part problem." Experimenters in this area have tried to answer the question of whether subjects will learn a stated amount of material faster if given the entire amount at once (the whole method) or if given the material in smaller blocks (the part method). Variations on the part method have also been tried. Woodworth and Schlosberg (1954) describe the "progressive parts" method in which the subject learns part I, then learns parts I and II together, and so on. In the "pure part" method, subjects learn part I, learn part II, and then put them together. (There may be more than two parts, of course. In fact, one of the questions asked in such research is "What size should the parts be, for any given amount of material to be learned?")

This approach to learning has had its problems. Underwood, writing in 1949, commented that little work had been done on the problem in ten years. He suggested that this lack of interest in what had once been a "hot issue" was due to two factors:

1. Results of whole-part studies were contradictory for no apparent reason.
2. There was not an adequate theory to direct research.

A review of whole-part studies by McGeoch (1931a) certainly bears out Underwood's first point. McGeoch compared most of the studies done before 1931 on a number of points, such as materials used, number of subjects, learning methods, and results. Her findings for paired-associate whole-part learning are of interest. She reports on studies by Ephrussi (1904), Neumann (1907), Pechstein (1918, 1921, 1926), and Brown (1924). All of these studies appear in reviews of whole-part learning, but McGeoch's figures reveal the fact that, except for the study by Brown, these experimenters used groups of subjects whose numbers ranged from one to nine, with a mean of about four. The findings of these experimenters are split about fifty-fifty between whole and part methods (superior), and no particular trend in materials or criteria used is observable, except that the whole method is generally superior if retention is the test rather than learning speed. Hsaio (1951) also found that the whole method is reliably better than the part method as the amount of time subjects must retain the material increases.

Orbison (1944) hypothesized that an increase in the length of the list of materials to be learned involves a disproportionate increase in interference. He gave his subjects lists containing 8, 12, 16, and 24 pairs. He found that at no length was the whole method superior, but that the part method became increasingly superior as the lists lengthened. His figures may be of interest to us (Table 1).

Table 1

Mean Number of Presentations Per Pair Required to Reach a Criterion of One Perfect Trial (Orbison, 1944)

No. of Pairs	Whole Method	Part Method	Difference
8	18.75	18.67	0.08
12	20.58	16.96	3.62
16	28.75	23.42	5.53
24	38.50	25.21	13.29

Two facts should be noted about Orbison's study. First of all, in the learning of a paired-associate list, the subject learns some pairs earlier than others. Unless pairs already learned are dropped from the list, learning time to one perfect trial is increased by the repetition of these pairs. Second, Orbison's data shows a large practice effect across the segments of lists learned by the part method. This practice effect may well be a factor in making the part method superior. It is possible, however, to give subjects practice prior to the learning of a long list. The same effect might be obtained with less effort. Orbison's longest list contained 24 pairs. If lists had been longer, and the experiment had been carried further over time, it is possible that this practice effect would have leveled off and the superiority of the part method might have been reduced. These possibilities must be taken into account. Lakenan (1913) made just such an observation. His data showed that the part method was initially better but the whole method improves after subjects have had some practice. Seibert (1932) found the whole method superior too, but her longest list had 12 pairs. She tested subjects after 50 minutes and again after 2 days. Her retention figures indicate almost no loss for either method.

Brown's figures (1924) show that the whole method improves with practice. He gave subjects two lists. One group learned the first list by the part method and the second, by the whole method, and the other group learned them in the reverse order. For the group that learned by the whole method first, the whole method was superior to the part method by 24%. When the whole method of learning was preceded by the part method, the whole method was superior by 75%. We have already commented on the statistically invalid size of sample for studies by Neumann and Ephrussi, but we could add, for what it is worth, that both found that the whole method improved with practice. Davis and Meenes (1932) found the whole method superior for learning of vocabulary materials in comparison with the part method, even before much practice effect could build up.

"Whole" and "part" in the studies cited, refer to finite, usually fairly short lists of material. Most of the "whole" lists were made up of 30 pairs or fewer. It is important to ask, at this point, what these findings mean for educators. Even when we seek to teach a finite number of pairs, the number is more often in the "hundreds" than in the "tens." It is entirely possible that the observed functions would change when the amounts of material referred to as "whole" and "part" were made substantially larger. For educational purposes, the segments are operationally "parts." We have already noted that, with practice, subjects can handle larger blocks of material. Cook (1937) suggested that when the unit-size is smaller than the optimum, the whole method is best, and when the unit-size is larger than optimum, the part method is best. What we need to know, then, is the size of the optimal unit. We also need to know whether practice effects level off after a point, and we need to know how much improvement in learning and retention will occur, for a given subject, between his first list and the point where his learning-to-learn levels off.

There is very little hard data available on these issues. Suppes (1963, 1964) has produced both data and a theory about size of segments. He points out that when learning is faster than forgetting, the block size should be as large as possible, and when learning is slower than forgetting, the block size should be as small as possible. He bases his functions on the idea that each exposure of pair X increases the probability of a correct anticipation of response X on the next trial, and presentation of other pairs causes forgetting (to some degree) of pair X, thereby decreasing the probability that response X will be given on the next trial. Clearly, the number of pairs in the list, the number of pairs presented between any two occurrences of pair X, will influence the probability of a correct response. Such other factors as difficulty of pairs in the list, interlist, and intralist interference due to similarity, and ability of the learner will also influence these probabilities. Suppes' formulation of his theory may prove very useful in designing research on the optimal size of blocks. His studies provide some encouraging data too. He gave subjects English-Russian vocabulary lists organized in blocks of 6, 18, 36, and 108 pairs (all subjects eventually learned 108 pairs). Each item was given 21 reinforcements or trials. The group that received all 108 pairs in one list performed best. Suppes reports that he is now working with blocks of 108 and 216 pairs and the blocks of 216 pairs are being learned faster than those containing 108 pairs. Suppes' descriptions of his derivations refer to the increments and decrements in probability as α and β . He does not enlighten us as to precisely how to calculate these two parameters. List length, difficulty of pairs and difficulty due to list similarities may all have to be included, and we do not know precisely how, at this point. Nevertheless, his equations may provide the second requirement mentioned by Underwood (1949, see above) -- that of a theory to direct research in the whole-part problem. This lack of theory that Underwood points out, is very much the issue here. Whole-part learning experiments were, for the most part, a matter of trial and error. The concept of list length, or amount of material to be learned, as a continuum is missing. It seems clear, from the whole-part research (if we can make any generalizations from it) that neither method is, per se, better. Perhaps it is not a question of method, but of the amount of material we wish the student to master. If there is an optimum "block size" for any given student, then giving him material in smaller doses is inefficient and larger amounts may only confuse him--thereby causing him to learn all the items more slowly than he might have otherwise. Suppes' work is helpful, but to a certain extent, he still belongs in the mainstream of whole-part research in that he is using arbitrarily selected block sizes and comparing methods. He overcomes this shortcoming in theory, but his theory is not as yet worked out well enough to be useful.

Interference Theory

It seems clear, from the results of whole-part studies and others, that, if we wish to optimize our students' learning, we will have to work with the variable list length, or amount of material per lesson, as a continuum. We might wish to obtain data on the learning of several different list lengths over time, that is, data from students who have received stated amounts of material to be learned each lesson over a period of time. But, as we noted earlier, it is possible that as the amount of material the student has already seen builds up, the student will find it harder to learn the given amount of new material. It may be that, as the student progresses through the lessons, his learning speed will decrease as a function of interference. If we may expect some interference to occur when students learn in the manner of our hypothetical student, then we need to know what conditions produce interference and what can be done to minimize it; we need to know what variables are involved and how they interact.

The two major types of interference to be considered, are "retroactive inhibition" (where "interference" and "inhibition" are interchangeable terms), and "proactive inhibition." Retroactive inhibition has most often been studied in the classic paradigm symbolized as A-B-A, where subjects learn list A, learn list B, and then recall and relearn list A. Assessment of inhibition is made by comparing these subject's scores on the relearning of list A with scores of subjects who did not learn list B, and the difference in these scores, when the experimental group performs less adequately than the control group, is evidence of inhibition. Proactive inhibition experiments most frequently use the paradigm A-B-B, where subjects learn list A, learn list B, and then relearn list B. Their scores are compared with scores of subjects who did not learn list A. Retroactive inhibition has received more attention generally, than proactive inhibition. Furthermore, most experimenters in the field have worked with lists designed to provoke maximum interference, lists in which the stimuli are identical and only the responses differ. This condition is often referred to as the A-B, A-C paradigm (where A refers to the stimulus set and B and C to the response sets). The symbols for the general paradigm and those used for the list paradigms are occasionally confusing. With these terms made clear, we can ask questions of the research in interference that may shed some light on our hypothetical student's dilemma.

1. What mechanisms operate in interference? What, for example, has happened to first-list associations when the second list is learned and interference effects show up on relearning trials?

Barnes and Underwood (1959) attempted to find evidence which would support either of two theories about the "fate" of first-list associations. Using the A-B, A-C paradigm, they tested results by a method they called the Modified Free Recall, or MFR. The subject is given a sheet of paper with the stimuli printed on it and told to write down either or both of the responses to these stimuli, or, failing that, the subject may write down any response he wants to. Subjects were stopped at various points in the learning of the second list and given the MFR. It was also administered just before relearning of the first list. Barnes and Underwood hypothesized that if, as learning of list-two progressed, more C responses were given and fewer B responses appeared (subjects were free to write both responses), then this would be evidence in favor of the "unlearning" or extinction hypothesis, which holds that as list-two responses are learned, list-one responses are extinguished and are no longer available to the subject. If, on the other hand, list-one responses continued to appear as list-two responses were learned, this would favor the hypothesis that both sets of responses are available to the subject and response dominance determines which one is given first. Further evidence for this explanation would lie in the fact that, if the dominance theory is correct, then the subject's task amounts to differentiating between the two sets of responses, and interference should show up early in the learning of list two, when subjects were attempting to distinguish or differentiate the two sets of responses. The experimenters gave the recall MFR soon after the learning of list two and found that responses from list one were significantly fewer than would be expected if the response set were available to the subjects. They voted in favor of the extinction hypothesis.

Briggs (1954) used the MFR in a similar experiment, also designed to test the extinction hypothesis. His subjects learned a practice list and list one (12 pairs of adjectives) on the first day. On the second day they learned list two, and were stopped at 3, 6, and 9 pairs learned and at one perfect trial to take the MFR. In this way, the experimenters hoped to assess the waxing and waning of response strengths. Subjects then relearned list one at 4 minutes, 6, 24, 48, and 72 hours after learning list two. Again they were given MFR's before relearning. (Subjects had also been given MFR's before the experiment began, to find out what associations they brought to the experiment for each stimulus.) Briggs' figures (Table 2) for the number of responses from each list given just before relearning are interesting:

Table 2

Number of Correct Responses Per List (Briggs, 1954)

	4 min.	6 hrs.	24 hrs.	48 hrs.	72 hrs.
List one	2.56	4.36	5.32	4.96	5.04
List two	8.20	5.20	4.80	4.08	3.80

He interprets these results as favoring the extinction theory. The negatively accelerated, positively sloped curves for list-one responses, compared to list-two responses, are exactly like the curves for the responses subjects brought to the experiment in comparison to the list-one responses. List-two responses decrease significantly between 4 minutes and 6 hours, with a gradual leveling off after that. List-one responses increase significantly between 4 minutes and 6 hours and reach a maximum at 24 hours, which they maintain over 72 hours. Briggs concludes that this shows a general acquisition and extinction function. Both Underwood (1958) and Briggs, Thompson, and Brogden (1954) tried to find out what goes on in those first 6 hours by repeating the experiment with relearning at 4, 12, 36, 108, 234, and 360 minutes after list two. Underwood found no significant increase in responses from list one. Briggs and his associates did. They explain this seeming contradiction by the fact that Briggs used a 24 hour lapse between list one and list two and used naive subjects, while Underwood ran the entire experiment in one day and used practiced subjects.

If the unlearning hypothesis is correct, if subjects must extinguish responses learned previously in order to pick up new ones, then interference, even when stimuli are not identical, may be a serious impediment to the optimization of students' learning in the classroom. If the subjects' behavior could be explained in terms of list differentiation and response dominance, on the other hand, these variables could be manipulated to provide the least possible interference for learners. It is worth noting in this context that the figures above, from Briggs' study, show that whatever unlearning took place, subjects had recovered much of their first list response set 6 hours later. The MFR's were given before any relearning of list one took place, so whatever subjects knew about list one at that point, they gained in the first learning of list one. It would appear that "unlearning" is no more than the temporary effect of list two on list one, as measured 4 minutes after the learning of list two. Considered in this light, unlearning, or whatever one chooses to call the observed effect, appears to be a temporary phenomenon from which subjects do recover spontaneously. (For a discussion of this issue see Postman, in Cofer (1961).)

An examination of the data reported above and data from other studies suggests two variables which might be maniuplated: the degree of learning for list one and list two, and the amount of time between the two lists and between list two and relearning.

McGeoch (1929) gave his subjects 6 to 26 trials on list one and 11 trials on list two. He found that the higher the degree of original learning (list one), the less retroactive inhibition was observed at recall. He attributed this to increasing list differentiation by subjects, as the strengths of the two response sets were differentiated by practice. Melton and Irwin (1940) found a curvilinear relationship between the degree of interpolated learning (list two) and the amount of interference observed when original learning was held constant. The maximum amount of interference occurred when list one and list two were equally well learned. The findings of Thune and Underwood (1943) show that, with original learning held constant, retroactive interference increased up to 16 trials on list two (compared to list-two results of subjects who did not receive list one) but 10 more trials yielded no more interference.

Proactive inhibition experiments with the same variables have been fewer. Waters, (1942) used lists of 18 consonant syllables and gave list one 1, 3, 5, 10, 20, or 40 trials. He gave list two 5 trials. His subjects recalled list two after 20 minutes and no relationship between number of trials on list one and performance at recall was noticed. Very little inhibition was produced. He points out, however, that his two sets of responses were of low similarity. Underwood (1949) demonstrated that list differentiation increases with differences in response strength and decreases as the time lapses between learning and recall lengthen. He gave subjects three degrees of learning for list one, held degree of learning constant for list two, and tested recall after 20 minutes and 75 minutes (different groups of subjects). He found significant inhibition in the learning of list two, but only on early trials (while subjects were presumably differentiating lists), and found significant amounts of inhibition at recall when recall took place 20 minutes after learning. Only the highest degree of learning for list one produced any interference in the relearning of list two after 75 minutes. Underwood points out, in this article, that these variables, time to recall and degree of learning, behave alike for both kinds of inhibition.

Briggs (1957) varied the degree of learning for both list one and list two. He again used the MFR test just before relearning (in a retroactive inhibition experiment). A given subject received 4 degrees of original learning and one degree of interpolated learning, so that there were 4 groups. His results show that, in general, an increase in the degree of original learning brings an increase in performance at recall of list one. His curves for responses on the first trial of relearning (the recall test) are shaped the same way for all degrees of learning on list two (curves across original learning for each degree of interpolated learning), but the absolute number of correct responses at recall increases with degree of interpolated learning too.

Relearning, however, is a different story--in Briggs' study and in others. In Briggs' study, mentioned above, the differences in magnitude on the first trial of relearning disappear by the third trial. Subjects who received 2 trials on list one performed as well as subjects who received 20 trials, for any given level of interpolated learning. Interference caused by interpolated lists learned to 20 trials dissipates more rapidly for lists originally learned to 2, 5, or 10 trials than does the interference caused by second lists learned to 10 trials.

Evidently, anything that serves to render the response strengths of the two lists unequal helps to prevent inhibition. In the retroactive inhibition paradigm, the fate of list-two responses after the relearning of list one is not generally assessed. McGeoch's finding, that a high degree of original learning decreases the detrimental effect of list two on recall of list one, does not tell us what state list two was left in after relearning of list one.

Briggs (1957) suggested, in his conclusion, that list differentiation is a major variable in the study of interference. His work suggests that the measurement technique used in such studies may also be of importance, particularly in studies of the problems of optimization of language learning. In Briggs' study, in the study by Underwood (1949) mentioned earlier, in the work of Newton and Wickens (1956) and many other studies, the effects of interference observed at recall are no longer present after a small amount of relearning. For the purposes of these experimenters, recall is a sensitive test. If we wish to work with the problems involved in language learning, however, it would be a mistake to say that recall and relearning are correlated (as Luh, in 1922, showed) and use only one of these measures. Relearning, it would appear, gives a better estimate of what the subject knows. Both recall and relearning scores would be valuable in the sort of optimization study we would have to perform if we wish to answer the questions posed earlier.

2. How much interference is due to similarity of stimuli and responses?

Studies using the A-B, A-C paradigm offer insight into the behavior of interference, but our hypothetical student's lists will, presumably, contain pairs in which the stimuli and responses are both different. Several experimenters have attempted to find the gradient for interference across similarity of stimuli and responses. Twedt and Underwood (1959) assessed the effects of mixed lists on the learning of list two. All subjects learned list one (A-B) to a criterion of one perfect trial. Four groups then received a second list composed of pairs in the A-C, C-B, C-D, or A-Br paradigms (A-Br pairs were stimuli and responses taken from list one but re-paired). Four more groups received second lists in which some of the pairs were assigned to each paradigm. There were no differences of significance between lists of each paradigm and pairs of each paradigm type in the mixed lists. All four paradigms produced negative effects during the first 10 trials of list two; the A-B, A-Br paradigm produced the most interference and the A-B, C-B paradigm produced the least interference. We could have hoped that the A-B, C-D paradigm would show the least interference, since it parallels the student's learning, but evidently similarity of responses is helpful. The A-B, C-D paradigm was second only to this response-similar one.

Osgood (1949) stated three laws, results of his work with similarity of stimuli and responses:

1. Where stimuli are varied and responses are functionally identical, positive transfer and retroactive facilitation are obtained, the magnitude of both increasing as similarity among stimulus members increases.
2. Where stimuli are functionally identical and responses are varied, negative transfer and retroactive interference are obtained, the magnitude of both decreasing as similarity between the responses increases.
3. When both stimulus and response members are simultaneously varied, negative transfer and retroactive interference are obtained, the magnitude of both increasing as the stimulus similarity increases.

Bugelski and Cadwallader (1956) designed an experiment to test these laws. Their stimuli were visual and categorized as: identical (I), very similar (VS), low similarity (LS), and neutral (N). Their responses were words from Osgood's lists and were divided into the same categories on the basis of Osgood's work with them. This yields 16 conditions when each category of stimulus is paired with each category of response. All subjects were run to one perfect trial on both lists. Their results showed that:

1. Groups with identical stimuli performed most poorly on recall tests.
2. Groups with identical responses performed best on recall tests.
3. Groups with neutral stimuli were in the intermediate position for performance on recall.
4. The group having neutral stimuli and neutral responses performed nearly as well as control groups on recall. Scores for this group were well above the scores for groups with identical stimuli. As Osgood predicted, for groups with neutral stimuli, the interpolated learning had little effect on recall or relearning.
5. Groups with similar responses performed more poorly than did groups with identical or neutral responses.

Bugelski and Cadwallader used their results to correct Osgood's second law, to read, "...the magnitude of both first increasing and then decreasing as similarity between the responses increases."

This study provides us with some interesting figures. Most of the studies reviewed here used the classic A-B, A-C paradigm. Bugelski and Cadwallader used it too, but they used an A-B, C-D paradigm as well. The comparison of the two paradigms is shown in Table 3.

Table 3

Results of Two Paradigms (Bugelski and Cadwallader, 1956)

	OL (trials)	IL (trials)	Recall (pairs)	Relearning (trials)
A-B, A-C	8.11	6.44	7.00	5.11
A-B, C-D	8.33	5.89	9.71	4.33
Control	7.67	____	10.56	2.67

The experimenters adjusted scores on recall and relearning for original learning, so that the individual subject's ability would not bias the results. According to their report, the differences between the above paradigms for recall and relearning are significant. The A-B, C-D paradigm ranked 7th out of the 16 paradigms used; the A-B, A-C paradigm ranked 16th. Moreover, their rankings hold for relearning times as well as recall scores. Their stimuli were visual figures, which renders the study less comparable to many school learning situations; and they dropped pairs from the lists after two correct anticipations. Even so, their findings are encouraging. Some interference did appear in the A-B, C-D paradigm, but not so much as appeared in the classic A-B, A-C paradigm. And, from what we have seen, the interference that did occur may well be a result of nearly equal response strengths, since subjects learned to the same criterion for both lists. It would be interesting to see what would happen if, using the A-B, C-D paradigm, the degrees of learning for the first and second lists were kept imbalanced by various degrees.

Morgan and Underwood (1950) tested Osgood's second law for response similarity in a proactive interference experiment. They used verbal materials (as opposed to Bugelski and Cadwallader's visual stimuli) and did not drop pairs as they were learned. Their subjects served in all of five conditions of response similarity (stimuli were identical). They found a decided practice effect in list-two learning, when experimental groups were compared to control groups who did not receive list one. They also found that rate of learning is directly related to response similarity; the more alike responses are (from list one to list two) the faster they are learned. All their learning curves for list two are similar in shape across degrees of response similarity, but do not overlap in magnitude. Their second list was learned only to a criterion of 7 out of 12 pairs correct (list one was learned to one perfect trial). At recall, 20 minutes after learning list two, the groups displayed a linear relationship between amount known and similarity of responses. It is possible, therefore, that the results of Bugelski and Cadwallader, which led them to rewrite Osgood's second law, are artifacts produced by the equality of learning, the dropping of pairs from lists as they were learned, or the use of visual stimuli with verbal responses. Or, it is possible that the proactive design reveals a variation of the learning mechanisms found in the retroactive design.

3. What are the similarities and differences in the behavior of proactive and retroactive interference? One of the questions we asked about the language student's learning procedures was whether he should be given old items for practice or new items for learning first in any one day. It would be to the student's advantage to learn last, that list which is most capable of causing interference and least likely to suffer from interference due to other lists. Youtz (1941), in attempting to validate Jost's laws (Hovland, 1951), confirmed the fact that older associations show larger learning increments after one relearning trial, and require fewer trials to relearn than younger associations. We could deduce from these and other findings cited earlier, that, in general, a list that is better learned is less likely to suffer from interference due to other lists. We could further deduce that such a well-learned list is more capable of causing proactive inhibition than is a less well-learned one. If this is the case, then we would most likely give our language student his lists in the order of A, B A, C B A, and so forth. List A is the best learned list. It should, therefore, suffer least from the effects of B and C, and it is possibly the list most capable of causing interference if relearned before B or C. We would have, then, a basically retroactive paradigm, allowing for multiple lists (an allowance which is by no means guaranteed not to change all results found so far).

Underwood (1945) summarized findings about proactive inhibition.

He noted that:

1. When stimuli are identical for both lists, amount of proactive inhibition increases directly with the number of lists learned prior to the list relearned. Retroactive inhibition shows this same relationship with reference to the number of lists interpolated and the amount of interference noticed at recall. But the effects of retroactive inhibition are greater than those of proactive inhibition.
2. Retroactive inhibition is greater than proactive inhibition for cases where recall and relearning took place shortly after the learning of the second list. After an interval of 48 hours between learning and recall there is little difference between the two types of interference.

Our student's situation is complex; we have not only to consider the effects of ways of ordering lists within any one day, but also the effects of overnight and weekend time lapses. We must further consider, if the evidence cited here is correct, the degree to which he should learn each list. It is difficult, to say the least, to choose between the Scylla of retroactive and Charybdis of proactive inhibition.

4. Does interference from sheer mass of material already presented build up over time? Most of the studies reviewed here

used only two lists. A few experimenters gave the same subjects more than two lists, but any interference due to amount of material previously learned was confounded by the type of lists used--stimulus similarity or identity would, as we have seen, cause a great deal of trouble by itself, and results were analyzed in these terms. No figures were reported for interference due to amount of material in these studies.

We could ask, first of all, what would happen when a student learns a given list of material on the first day, and then relearns that material on subsequent days. How much time would he save each day? Is there any constant ratio between any one day's relearning time and the learning time for the day immediately preceding it? The early work of Ebbinghaus (1885) would suggest that there is such a ratio. As can be seen in the figure below (Figure 1), the amount of time needed to relearn on the second day is approximately two-thirds of the first day's time. The ratio seems to hold up across the six days shown. Ebbinghaus used serial lists (8 of them) and, of course, an N of one, but it is not unreasonable to hope that some such ratio exists for paired-associate learning too. Given this ratio, we could make fairly accurate predictions about the amount of time students need to learn and relearn a given list to a stated criterion.

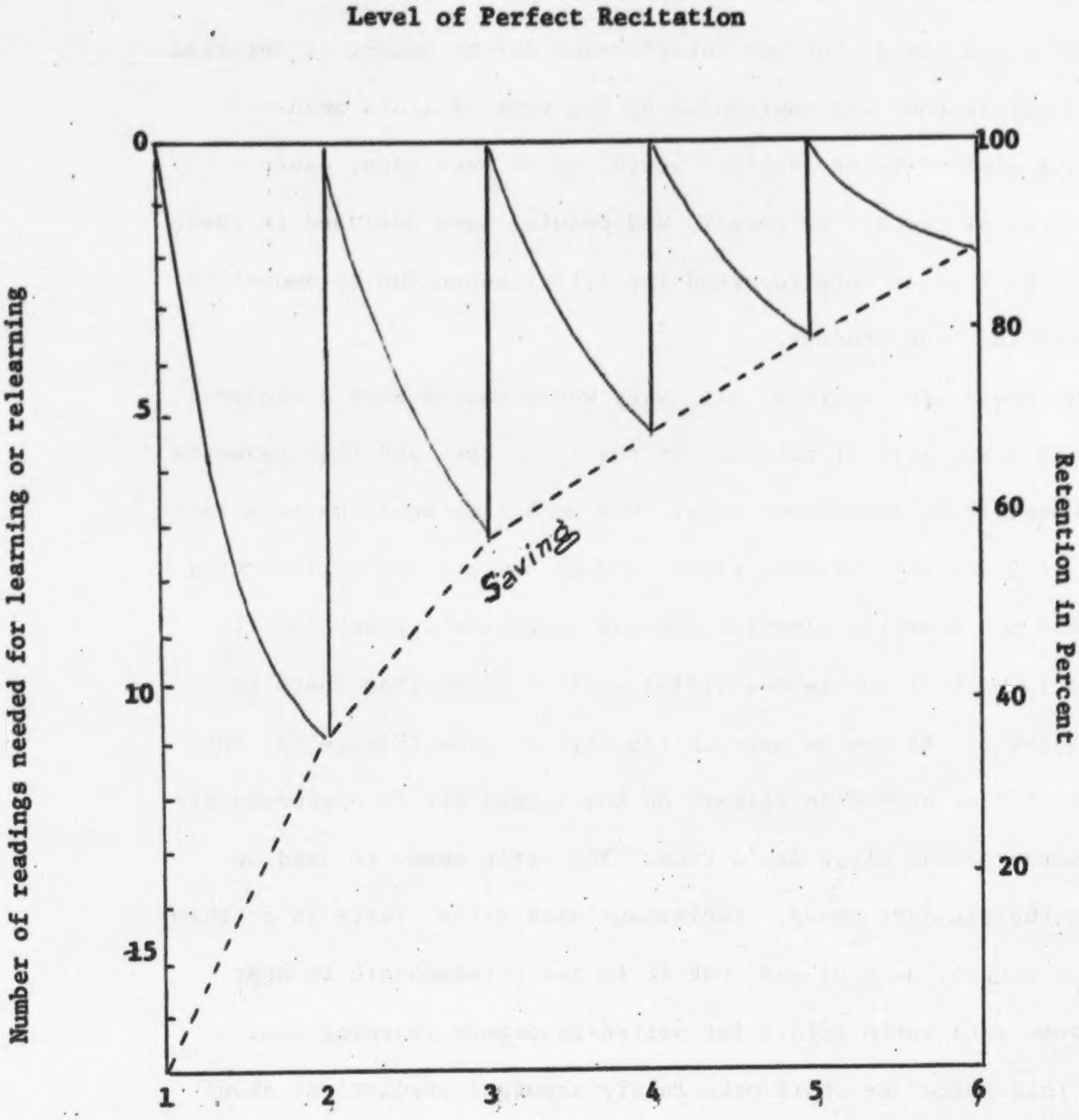


Figure 1. Data from Ebbinghaus (Woodworth and Schlosberg, 1955, p. 730).

But what would happen to that ratio if we gave the students a new list to learn each day, and required them to relearn all past material until they could recall it satisfactorily? At what point might interference effects "snowball" and cause our students to slow down so that they could not learn a satisfactory amount of material in the hour or so alloted to language study?

Underwood and Schulz (1961), in a study of massed and spaced practice, gave their subjects four lists of eight syllables. ("Spaced" practice includes an intertrial interval and "massed" practice does not.) Learning took place across four days. Subjects learned list one on the first day (by either massed or spaced practice), learned lists two and three on the second day (all by massed practice) and also relearned list one for five trials, and on the third day, learned list four by either massed or spaced practice. The fourth day was devoted to five trials of relearning for list four. They found interference for all lists. Distributed practice produced poorer performance on list four than did massed practice. Our concern with their results lies in the figures for learning and relearning of lists one and four (the only ones reported). For lists in which the stimuli were English adjectives and the responses were medium association-value nonsense syllables, the percentage of items recalled are:

List one	78%	List four	36%

Both lists were given 12 learning trials and tested the following day (list one was tested on the second day before any new learning took

place). The difference between these two percentages is presumably due to the learning of lists two and three on the second day. The experimenters used lists with stimulus similarity across lists (not stimulus identity) which may well account for a lot of the interference noticed.

Underwood and Richardson (1955) performed an experiment with basically the same design. The first and last lists were tested for recall and relearning, and were learned by both massed and distributed practice. Two groups learned two lists (no interference), two groups learned four lists, and two groups learned seven lists. Learning took place on four days--at least for the groups receiving seven lists. Table 4 illustrates the schedule.

Table 4

Learning Schedules (Underwood and Richardson, 1955)

	Two-list groups	**Four-list groups**	**Seven-list groups**
Day 1	**1 list**	**1 list**	**1 list**
Day 2	**1 list**	**2 lists**	**2 lists**
Day 3	**relearning**	**1 list**	**3 lists**
Day 4		**relearning**	**1 list**
Day 5			**relearning**

All lists except the first and last were learned by massed practice for both groups. List one was relearned on the second day, before

a new list was presented. All lists contained four pairs of nonsense syllables. Table 5 shows the mean number of correct responses at recall of the last list, for both massed and distributed groups. The authors conclude that retention decreased consistently as a function of the number of lists presented before the list being tested, but not as a function of intertrial interval, or type of practice. Their lists contained only four pairs each, and this brevity may have masked an even more massive interference effect.

Table 5

Mean No. of Correct Responses at Recall (Underwood and Richardson, 1955)

List Preceded By:	Mean No. Recalled (out of four)	
	Massed	Distributed
one list	1.25	1.46
three lists	.82	1.11
six lists	..75	.73

Interference theorists have not often worked with multiple lists. On the basis of the few studies cited here, we can only say cautiously, that it looks as if interference does increase as the number of lists learned increases, bearing in mind the similarity of Underwood and Schulz's syllables and the brevity of Underwood and Richardson's lists. We will have to look elsewhere for more data.

Ability

How much does the subject's ability affect his learning of paired-associate material? What sorts of abilities does such learning require? What instruments will assess such abilities so that accurate predictions of learning times and retention can be made? One of the claims made for programmed instruction is that it can tailor the instruction and the pacing to the needs of each student. But such tailoring requires precise knowledge about the abilities involved in any particular type of learning and tests which measure these abilities accurately.

Cook (1936) found that individuals differ widely in the amount of material they can handle in one block, but that this "optimal" block size improves with practice. McGeoch (1931b) asked if the relative efficiency of whole and part learning methods was a function of subjects' IQ's. She used children (ages 9 to 11 years) as subjects. For English-Turkish pairs, there were no differences between children with IQ's above 140 (Binet-Simon test) and children with IQ's between 95 and 105. For English-nonsense syllable pairs, the slower children preferred the whole method, while the brighter children performed about equally well on whole learning and progressive-part learning. These findings held for recall and relearning too. Gillette (1936) offered a possible explanation of these findings. He demonstrated that when pairs were removed from a list upon being learned, fast learners performed considerably better than slow learners on retention tests. In an experiment where pairs were not dropped, this difference in retention did not appear. Gillette hypothesized that the slow learners got extra practice on already-learned pairs while they were trying to learn the last few pairs in the list. Since, by definition, slow learners receive more trials on a given list than do fast learners, differences between the two types will not show up in studies that do not drop pairs. It would appear, then, that we shall have to allow more time, more practice, to slower learners in the classroom.

Asher (1962) has provided a reassuring piece of evidence for the relevance of paired-associate research to language learning. He correlated results of the Modern Language Aptitude Test (Carroll and Sapon, 1958) with data from a paired-associate task given to the same subjects, and found that the entire test correlated .764 with the learning task. The paired-associate subtest of the MLAT correlated .707 with the same task. Asher points out that the MLAT predicts total language learning quite well, and that, since it, in turn, correlates highly with a paired-associate laboratory task, we can have confidence in such laboratory work as relevant to language learning. Kjeldergnard (1962) also found that the MLAT predicted such learning very well. His correlation between scores on Part V of the MLAT and the number of correct responses in trials 3 to 12 of a paired-associate task was .60 ($p = .01$). The correlation between Part V and three successive perfect trials was .40 ($p = .01$). He also suggested that shortening the amount of time allowed for

Part V by 25% would improve its predictive ability. Carroll and Burke (1965) also found Part V of the MLAT to be a successful predictor of learning time.

Gladis and Braun (1958) tested the hypothesis that age differences would lead to differences in amount of interference (in a classic interference paradigm). They gave their subjects vocabulary tests and a practice list, which was used to predict learning ability. They found that there were no differences in amounts of interference due to age after scores for recall and relearning had been adjusted for vocabulary level and learning ability. Carroll and Burke (1965) found that for pairs of medium difficulty (low-frequency words from the Thorndike-Lorge lists), slow and fast learners (categorized by MLAT scores) acquisition curves differed. They suggested that this difference might be correlated with vocabulary level, since the materials used were unfamiliar English words.

Meaningfulness

The term "meaningfulness" includes a multiplicity of measures; any characteristic of the members of the pairs used in a learning experiment which makes a difference in the subject's learning could be included here (for a review of this question, see Higa, 1961, pp. 17-20). In at least one sense, it is futile to speak to the foreign language teacher, for example, about one kind of material being easier to learn than another. If French is what he teaches, he cannot help the fact that something else might be easier to learn. But if the "meaningfulness" of materials affects subjects' learning and retention, the teacher can at least assess the level of his material and adjust his lessons accordingly. Further, if we wish to experiment with language learning, we must be prepared to make some statement about the generality of our findings over various levels of "meaningfulness." Carroll and Burke (in press) used three different levels of "meaningfulness" in their experiment. High-frequency words from the Thorndike-Lorge list made up the "high meaningfulness" lists; low-frequency words from the same list made up the "medium meaningfulness" lists; "low meaningfulness" materials were nonsense syllables of 0 to 35% association value (Archer, 1960).

Their experiment was a factorial one, with four list lengths, two levels of ability, and three rates of presentation in addition to the meaningfulness levels. They found a large "meaningfulness effect" but no interactions with any other variables, in the analysis of variance. Noble (1961a) gave subjects two lists, to see if meaningfulness (the term is his creation, as a matter of fact--see Noble, 1961b) facillitated learning of the second list. He found that, as "meaningfulness" or (m) values increase, learning is easier, but no transfer effects were noticed. Cieutat, Stockwell, and Noble (1958) found that differences in meaningfulness (in Noble's terms) were more important when the response member of pairs was varied than when stimulus members were varied. Again, however, the facillitation was straight-forward; as meaningfulness increased, learning was faster.

It would appear, then, that the level of "meaningfulness" makes a difference in subjects' performance in learning, but does not interact with most other variables. Its effect is linear and fairly predictable. Experiments in paired-associate learning could conceivably be conducted at one level of meaningfulness, and the effects of other levels of meaningfulness could possibly be inferred. This savings in time, subjects, and effort could very well prove valuable when other variables are of prime interest.

"Warm-up"

Newton and Wickens (1956), in a study of interference, asked if the usual A-B, A-C paradigm, which provides massive interference, might not obscure interference due to other variables. They gave their groups an A-B, C-D paradigm and gave the interpolated list immediately after list one (the 0 condition), 24 hours after list one, and 48 hours after list one. All relearning took place after 48 hours. They hypothesized that, in the 48-hour group, the interpolated list might cause interference but that it also served to "warm-up" the subjects. They therefore gave their other two groups a short "warm-up" just before relearning. Their hypothesis was correct; the 0 and 24-hour groups performed better in recall and relearning than did the group whose "warm-up" consisted of list two.

The 0 and 24-hour groups also performed better than control groups who received no warm-up. Irion (1948) showed that the loss of retention found in conjunction with longer time lapses between learning and recall may be due to lack of warm-up. One group of his subjects recalled after a minute. The other group recalled after five minutes, with a short warm-up on color-naming. The five-minute group performed better than ~~did~~ the one-minute group. Thune (1950) compared the effects of a warm-up using paired-associate lists of various lengths to that of a warm-up on color-naming. He found that both tasks improved the learning of list two. (He gave several numbers of trials on the warm-up lists and the facilitation effect increased with the number of trials on the warm-up lists.)

Surroundings

In seeking ways to minimize interference, Nagge (1935) tried giving subjects the original lists and the interpolated lists in different rooms. He found no significant effect. Bilodeau and Schlosberg (1951), however, varied the room, the location (attic or basement), the mood or tone of the atmosphere, the method of presentation (card presenter and memory drum), the subject's position (standing or sitting), and everything else they could think of to differentiate the original and interpolated lists. The two sets of surroundings were as different as it is possible to make them. They found distinctly less inhibition at recall and relearning of the original list. Greenspoon and Ranyard (1957) replicated this study with the addition of a comparison between practiced and naive subjects, and found that changing the surroundings does lessen interference for naive subjects but not for practiced ones. Pushed to its logical conclusion, this theory gives rise to visions of classes of language students wandering through schools in search of an untried classroom, but the findings are there, and perhaps worth considering, since they add evidence to the theory that list differentiation is of major importance.

Summary of Research Findings

The research discussed here does not offer direct answers to the six questions posed earlier. Whole-part learning experiments were, as we noted, generally not very well designed and not amenable to statistical analysis. Even the better ones do not offer much in the way of guidelines for the educator, since variation of the amount of material involved may well change the results, and there is no really satisfactory way to extrapolate a general, predictive equation from the different studies. But the research in whole-part learning does point to the need for studies of list length as a continuum, over time and over fairly large amounts of material.

Interference studies have several drawbacks too-primarily experimenters' predilection for studies with similar or identical stimuli and for studies involving only one or two lists. Even in the studies where more lists were used, the total amount of material was generally not large, and not all lists were tested for recall or relearning.

These remarks are not meant to be critical of experimenters in paired-associates; indeed, anyone who studies the field will find the quality of the research admirable (particularly the immense number of excellent studies by Underwood and his numerous associates). But it will clearly be necessary to experiment further and in new directions, if we wish to answer the six questions we posed earlier. And, if we have not found answers to the questions, we may find that past research offers much in the way of ideas for designing new experiments. If we consider the plight of our hypothetical student, as he faces new material every day and is required to remember all that he has learned, we might assume, first of all, that we would wish to minimize interference for him, as much as possible. The questions we seek to answer assume that no unnecessary sources of interference such as similarity of material or an inefficient ordering of material are present. It would be worthwhile, then, to ask what we can do, judging from past research, to minimize interference. The following possibilities suggest themselves:

1. List differentiation by means of unequally learned lists. It would seem to be best to let students learn lists to different criteria, or to present lists in such a way as to keep the degree of learning different for each list.
2. Ordering of lists so that the best-learned lists are

practiced after less well-learned lists. Evidence from proactive and retroactive studies, although not conclusive, would suggest that better-learned lists are more capable of causing interference and less likely to suffer from it. The conditions under which our subjects would learn have not been studied extensively; it is possible that this "maxim" will not hold true. But it seems to be the case for paired-associate learning so far.

3. Provision of enough time for earlier associations to "recover." It would be a waste of effort and misleading if we failed to allow enough time for subjects to recover their associations to earlier lists by testing too soon. The optimal times range from 24 to 72 hours; 24 hours seems to be enough time for subjects to stabilize their associations for different lists. Given that we would use different stimuli and responses in each list, this precaution may not be necessary, but it cannot hurt and may help.

We cannot predict now whether students' learning behaviors will show any effects due to interference, when they are asked to learn large amounts of material over time and tested for recall of all material. In studies where the similarity of stimuli and responses were varied, lists of dissimilar stimuli and dissimilar responses were learned with less interference than were lists containing similar stimuli. But some interference effects appeared; these lists were not learned so easily as were lists in which the responses were similar. However, these studies (notably Osgood's) did not test many lists over time, and subjects were not always given optimal amounts of time in which to "recover" older associations. Thus far, the mechanisms of interference seem to be largely a matter of similarity of materials and timing of the testing for recall, the arrangement of lists and other such factors. If we wish to know what would happen, what subjects would do when we have controlled these sources of interference as much as is possible, we shall have to experiment further.

CHAPTER II

Experimental Design

As we have noted, paired-associate learning in the educational setting usually has three conditions not always found in past experiments:

1. A constant amount of time is available per day for learning
2. Students are expected to remember all that they have learned while they are learning new material (this expectation is not always fulfilled under even the best of conditions, but theoretically, it exists).
3. The amount of material to be learned is large and the time set aside for learning a subject is long, usually a semester or a year. Most subject areas can be broken into segments somewhat shorter than the entire semester; however, the material still exceeds the usual few experimental lists.

Earlier we noted that, if we wished to meet these three conditions in such a way as to make optimal use of the student's time, we would need answers to at least six questions. The full answers to all six questions, together with all the subsidiary questions that are bound to arise, comprise a career, not an experiment. If we wish to stay within the bounds of reason, we might well ask, as our first experimental question "Where do we start?" That is, if optimization of learning under the above conditions is our aim, what is the best, the most productive way of working with the variables involved. What variables are involved in students' learning behaviors under our conditions? The entire experiment to be discussed here can be viewed as an attempt to begin to identify the variables involved in educational paired-associate learning, particularly those variables that must be considered if optimization of students' learning is to be achieved. If we can successfully identify some of the variables involved in such optimization of learning and their relationships to each other, so that we can begin to form predictive equations for learning behaviors under our conditions, we will have taken the first step toward answering our questions. Before we can begin to explain how students learn, how they do what they do, we must first know what they do or will do under given conditions. Past research has not offered much that would predict what students will do under the conditions of our experiments.

The first requisite for new experimentation, then, is the maniuplation of list length, or amount of material, in such a way as to provide at least three points on a continuum.

The second requisite of the design is that it provide information about interference that would useful in optimizing learning times. It may not be possible to begin by trying to optimize learning; we do not know enough to be able to predict whether students learning new material every day will show effects of

interference in their behavior or at what point in the learning sequence such interference will appear. It will be necessary, then, to find out how much interference builds up under several different conditions of list length, and hope that we can draw some conclusions or make some predictions about their behavior, such that further experimentation with optimization would be successful.

Thirdly, it will be necessary to let students learn enough material over a sufficient number of days so that interference, if it does occur under our conditions above, will begin to appear. We can only guess at the amount necessary to provoke such "cumulative interference."

Finally, it will be necessary to require that students practice or relearn all past lists while they are acquiring new ones. This "practice" need be no more than a single trial in which the student demonstrates that he still remembers the list; if he does not remember it, then he must be required to relearn it.

Some of the other variables mentioned in the first chapter can be included at little cost. Ability can be included by the use of tests prior to the experiment itself. Meaningfulness, or difficulty level, presents problems. If we were to include several levels of meaningfulness at each level of list length, etc., we would have an unmanageably large experiment on our hands. Moreover, the standard "meaningfulness" levels used in paired-associate research may not correspond to the difficulty level of any given subject area. So far as we know from past research, meaningfulness

does not interact in any complicated way with the variables we are investigating; a final check of this point will have to wait for another day. Further, the experiment to be described is intended as an example of the kind of research needed for optimization of learning; educators working in individual subject areas would have to adapt the method as well as the conclusions in planning curricula or programs.

With these considerations in mind, the following design was created and carried out.

Experimental Design

The experimental design is actually a series of experiments arranged in a systematic progression. There were nine groups in the experiment, and they were arranged under the various conditions described in Table 6. The same basic procedure was used for all groups. Each subject learned a list of paired-associates on Day One, and relearned that list on the second, third, and fourth days. On each succeeding day after Day One, a new list was presented. Each of these lists was relearned on succeeding days too. New lists were presented before practice on old lists took place. Thus, on the fourth day, for example, the student received List Four, List Three, List Two, and List One in that order. The lists varied in length as indicated in Table 6. There were three list lengths used: 42-pair lists, 26-pair lists, and 16-pair lists. These list lengths are arbitrary; the shortest length permits some comparison with past research, and the longest list was chosen in order to tax subjects' learning abilities, so that we would be reasonably sure to find interference if it exists when the amount of material is large but does not appear for shorter lists. The three list lengths are approximately equidistant on a logarithmic scale; Carroll and Burke (1965) found that list length behaves linearly on a logarithmic scale. The control groups learned only one list, on Day One, and relearned that list on succeeding days without receiving any new lists. Their learning times were presumed to be free of any interference effects that might be due to the presence of other lists. There was a control group for each list length. The basic design includes list length at three points on a hypothesized continuum and it includes the requirement that students remember what they have learned. It also includes a recall and relearning measure; all students returned after 72 hours to recall and relearn all lists in the same order as they had been presented on the last of the four learning days. The design makes no attempt to optimize learning; that cannot be done until we know more about what happens under the various conditions of list length and interference to be studied here. It was hoped that results from the experiments would lead either to some tentative conclusions about the best way to optimize learning, some parameters for such a procedure, or at least to a more organized experimental attack on the problem by pointing up what needs to be studied among the possible variables involved in optimization.

Table 6

Schema of the Experiment

Initial List Length

Amount of Interpolated Material	Short	Medium	Long
Control Groups	$16_{\underline{00}}$	$26_{\underline{00}}$	$42_{\underline{00}}$
16 Pairs Interp.	$16_{\underline{16}}$	$26_{\underline{16}}$	$42_{\underline{16}}$
26 Pairs Interp.		$26_{\underline{26}}$	$42_{\underline{26}}$
42 Pairs Interp.			$42_{\underline{42}}$

N = No. of pairs in List 1 (given on Day 1 and relearned every day thereafter)

Subscript (underlined) No. of pairs in all other lists (given on Days 2, 3, and 4, and relearned every day thereafter)

Experiment I. The control groups form the basis for the first set of hypotheses or, more correctly in this case, questions. These are the three groups which will learn only one list. Their times for learning and relearning are presumed to be free of interference due to added material. If we recall the curves shown in Ebbinghaus's data, we would hypothesize that their relearning times would follow some regular, discoverable pattern and could be predicted. We would ask, then:

1. What kind of learning curve is characteristic of the control groups?
2. Do learning curves differ for the three different list lengths and if so, in what way?

Experiment II.

The second set of questions or hypotheses involve the groups in the lower diagonal of Table 6, the groups which received the same amount of material on all four days (16-16, 26-26, and 42-42). If the control groups can be considered as the "best of all possible worlds" since they will receive no added material to cause interference in learning, then these three groups comprise the worst possible conditions. We are assuming here that the Ebbinghaus curves have some validity and that, if we wish to use no more time on succeeding days that we used on Day One, then there is no point at all to giving subjects more on succeeding days than they received on the first day. Relearning will have to occupy at least a small portion of the second, third, and fourth days. Later experiments may prove that it is feasible to increase the amount of material over days, but we do not now know what to expect in the way of interference and so will abide by the relationships suggested in the Ebbinghaus curves. The list lengths in these groups will most likely produce some interference, but "how much" and "what kind" we may well ask. Specifically, we wish to know:

1. Are there differences among the learning curves for the first list, List One, for the three groups?
2. Do the three initial list lengths (those given in the three broups on Day One) differ in learning times and relearning times from their respective control groups?
3. Do the learning and relearning curves for lists of the same length given on succeeding days differ from the list given on Day One? Since all of the lists given to any one group are of the same length, we can assess the differences in learning curves for the four lists that subjects will see.
4. Do these second, third, and fourth lists differ from the control lists for the respective groups? That is, do learning and relearning times for later lists show interference effects and if so, can we describe them?

Experiment III. The third set of questions to be considered utilizes the groups designated in Table 6 as having 16 pairs interpolated. These subjects all receive 16-pair lists on the second, third, and fourth days of the experiment. The groups differ only on their first lists. We are again assuming that optimization of learning will have to deal with interference and that it might be interesting to look at our three different initial list lengths when the interpolated material is held constant, and when the lengths of the interpolated lists are shorter than the length of the initial list, or at least equal to it, in the case of the "16-16" group. We could ask, then:

1. Do the learning and relearning curves for the three initial list lengths differ when the amount of interpolated material is held constant?
2. Do the learning curves for the three initial list lengths differ from those of their respective control groups?
3. Do the second, third, and fourth lists differ across the three groups, with respect to learning and relearning times?
4. Do the second, third, and fourth lists in each group differ from the control group list at the same length (16 pairs)?

Experiment IV. This experiment deals with the reverse of Experiment III. We ask what would happen when the initial list is held at one length and the amount of interpolated material is varied. Time and money place some limitations on the scope of this study, and it is not possible to ask this question about all three initial list lengths, but some information about the longest initial list length (42 pairs) can be gotten at the expense of one more group (group 42-26 in Table 6) since the other two groups have already been used in earlier experiments. Given three groups with the same initial list length and three different amounts of interpolated material, we can ask:

1. Do the learning and relearning curves for the 42-pair initial list differ across groups?

2. Do the initial lists in these three groups differ from the 42-pair control list with respect to learning and relearning curves?
3. How do learning and relearning curves for the second, third, and fourth lists in these groups compare with those of the control groups at equal lengths?
4. How do learning and relearning curves for the second, third, and fourth lists in these groups compare with lists of equal length given second, third, or fourth in other groups? That is, how do the 16-pair lists in group 42-16 compare with the 16 pair lists in groups 26-16 and 16-16? How do the 26-pair lists in group 42-26 compare with those in group 26-26 (all other conditions, such as day presented, being equal)?

Other Variables

Ability. We will wish to ask how ability is related to performance at every point in the experiment. That is:

1. How useful are current predictors of ability, particularly the MLAT, Part V, in predicting performance on the various lists in the experiment. The test will be given before the experiment to all subjects and correlated with each data point (each learning time for every list).
2. Is ability, as measured by the MLAT related equally well to all parts of the subject's task? That is, does it

predict learning and relearning times equally well?

Recall. One of the requirements we placed on experimentation with paired-associate learning in the educational setting was that the students must remember all that they learn, while they are learning new material. The first trial of each relearning can be considered a recall test and will be, perhaps, more sensitive to interference than relearning times will be. We have seen, from past research, that the two measures are not identical. The amount of time permitted between lists on any one day will be two minutes. Subjects will have 24 hours between learning sessions. A final check on recall and relearning will be made after 72 hours have elapsed, and this fifth session will not include any new lists to be learned. We can ask, then,

1. How do recall scores (as measured by the number of items the student correctly anticipates on the first trial of a list being relearned) compare in experimental groups and between experimental groups and control groups? We would obviously make such comparisons only where they were logical. Recall for lists of different lengths could be compared by expressing scores in percentages, although these are not always satisfactory.
2. How do recall scores for all lists in the final testing session (72 hours after the fourth learning day) compare with recall scores on the fourth day for each subject and group?

Clearly, subjects may not recall lists perfectly. But most educators aim for a satisfactory amount of recall, and, in discussing the merits of the various list lengths and combinations of list lengths, we will wish to compare recall scores; faster learning means little to the educator if the material is not well-remembered.

Errors. If we wish to analyze interference in the various experiments, we will want to ask:

1. Are most of the errors ones of omission or do subjects give responses that are incorrect? What ratio do subjects maintain between omissions and intrusions? Does the ratio change with list length or over days?
2. If intrusions occur, what portion of them are responses from the same list given to the wrong stimulus, what portion are responses from other lists, what portion are stimuli from other lists and what portion of the intrusions are items not on any list the subject has seen?

Intrusions from the same list, the list being learned, are common in paired-associate experiments. Intrusions from other lists might prove interesting, depending on how many of them are given, where they are given, and what list they come from.

Experimental Details

Subjects. It would be desirable to work with students in high schools, or, perhaps college students in classes, but this is not practical. The experiments required about two hours of time per day, over a week (estimating generously). No schools known to the experimenter would countenance the interruptions and missed classes that this time requirement would mean. Subjects were, therefore, Harvard and Radcliffe College undergraduates who responded to an advertisement and were paid for their services. They were naive subjects; no one who had participated in a verbal-learning experiment of a paired-associate nature was accepted. It was felt that naive subjects were a necessity since both Harvard and Radcliffe Colleges select students of high verbal ability, and the added and irrelevant facility of practiced subjects might have rendered subjects less comparable and might have masked differences in list lengths, etc. There were ten subjects in each group. That number is small, and more subjects would certainly have been desirable, but again time and money made small groups necessary. Subjects were paid a flat fee for their services, in order to discourage slower learning, which might have been the case if they were paid by the hour. The fee was adjusted for groups with longer lists to learn. (Subjects in the 42-42 group were paid fifteen dollars for the experiment, as were those in the 42-26 group. All others were paid ten dollars. The fee was based on a pre-experiment estimate of about ten hours work at a dollar and a half per hour.)

Three subjects were dropped and replaced because they failed to show up after beginning the experiment; two subjects were replaced because they could not learn in a reasonable amount of time.

Materials. All subjects received a practice list before each learning session. The list consisted of eight pairs; stimuli were letters of the alphabet and responses were two-digit numbers. Both stimuli and responses were chosen at random and paired at random, with the restrictions that, in the make-up of the list, no letters which precede or follow each other immediately in the alphabet were given in sequence and no two responses began with the same digit (as, for example, the numbers 92 and 97 would not both appear as responses). The eight pairs in the practice list are given in Appendix A.

Experimental pairs were made up of high frequency Thorndike-Lorge (1944) words as stimuli, and very low frequency, five-letter words from the Thorndike-Lorge List were used as responses. The stimuli were chosen from the lists in the "A" and "AA" categories. The responses were from the list of words which occurred at least 15 times, but less than 2 times per million on the Lorge magazine count and on the Thorndike general count. Most of the stimulus and response words had been used by Carroll and Burke (1965). The materials were chosen because it was felt that they represented the most reasonable compromise between making the materials too easy for the subjects and making them too difficult to be learned feasibly in a practical amount of time. The experiment required, at maximum, four lists of 42 pairs each, and it was felt that nonsense syllables or CVC trigrams would provide too difficult a discrimination task for subjects who might have to learn and practice as many as 42 x 4 of them in one day (the last learning day for group 42-42).

There were four forms of each list for any given list length. That is, there were four forms of the 16-pair list (Forms A, B, C, and D), and four for the 26-pair list, four for the 42-pair list. There was a possibility that subjects' learning curves over four days might be some function of the order in which the forms of the lists were given. Therefore, the forms of the lists were assigned in a systematically balanced manner. Since there were ten subjects in each group, Forms A, B, C, and D were assigned for the first day's list to two subjects each (a total of eight), and the remaining two subjects were assigned to one of the forms at random (with the restriction that both subjects did not receive the same form). The result was that Forms A, B, C, and D were split up among the subjects in the ratio 2-2-3-3, for the first day's list. On succeeding days, subjects in any given group received the other forms in a random order, with the restriction that any one form could not be too much used on a given day (day two, three, or four). This method worked for any length of list or combination of list lengths, as in groups 42-16, 42-26, 26-16, because the pairs for the 26-pair lists were selected from the 42-pair lists, and the

16-pair lists were selected from the 26 pair lists (all within a given form of the list, as 26-A was selected from 42-A, etc.). Forms were discrete; no pair in Form A appeared in Forms B, C, or D, for example.

The 42-pair lists in each of the four forms contained two responses beginning with the same letter of the alphabet, for a total of 40 pairs, and two stimuli beginning with letters that were not repeated (this composition of lists was due to the materials available--five-letter words at the level of frequency used were not inexhaustible). It was felt that the 26-pair and the 16-pair lists ought not to be totally free of alphabetic repetitions, since it was impossible to eliminate them from the 42-pair lists. Consequently, there were four pairs of responses beginning with the same letter in the 26-pair lists, and two pairs of responses beginning with the same letter in the 16-pair lists. It was felt that shorter lists made up entirely of such pairs of responses would be biased, since such similarity is more noticeable in shorter lists. Each form of a given list length contained the same number of responses beginning with a given letter. Insofar as was possible, the stimuli were also organized on this basis. Appendix A contains the stimuli and responses which were used.

Method. The stimuli and responses were presented to the subjects on a Lafayette Model 303B memory drum. The stimulus member of the pair was exposed for two seconds and the stimulus and response members together were exposed for two more seconds. There was no intertrial interval. Carroll and Burke (1965), in working with similar materials, found the four-second exposure slightly superior to both a three-second exposure and an eight-second exposure, and so it was used here. The anticipation method was used, and the subject was permitted to use any pronunciations he wished, so long as he was consistent.

Procedure. Subjects were run individually, in a room with only the subject and the experimenter present. All subjects received instructions which were read, in order to insure uniformity. All subjects were cautioned that their payment depended upon completion of the experiment, before any lists were given. Following the instructions, subjects were given Horton's adaptation of Part V of the MLAT (Carroll and Sapon, 1958). The shortened times suggested by Kjeldergaard (1962) were used. Subjects were then instructed in the use of the memory drum and given five trials on the practice list. (They were told that it was a practice list.) After the practice list, the first experimental list was given, and subjects were dismissed. Any questions about the aim of the experiment or the next day's task were pleasantly put aside with a remark to the effect that there was nothing secret about the experiment and everything was just what it appeared to be, but it was considered "bad form" for the experimenter to discuss it before it was over. Subjects were assured that full answers to any questions would be forthcoming after they had completed the experiment. On succeeding days, the practice list was again given for five trials, as a warmup. The day's new list was given precisely two minutes later, and the old lists were given, again at intervals of two minutes between lists. There were three "experimenters" administering lists; a given subject had the same experimenter and the same room throughout his sessions.

Subjects generally did not need many sessions before they knew what to anticipate on the following day. The experimenters parried questions, hopefully, with a manner that suggested that it did not matter how much the subject guessed and refused any comment on the grounds of "just not considered good experimental form." Subjects were not instructed to avoid rehearsing past lists, because it was felt that the surest way to insure that subjects rehearsed was to tell them not to. Subjects were instructed not to discuss specific pairs with friends, if they had any who were considering being subjects, since their friends very probably would receive different lists and would only be confused by such information. The experimenters generally knew of such "friends" and checked on their performances; any flagrant disregard for the request would have been noticeable, and no evidence of any "advance information" was discovered.

Criterion. The choice of a criterion for learning had to be made in such a way as to balance the possibility that our highly select population would learn too fast against the possibility that longer lists would not be learned in a practical amount of time (also the possibility that subjects presented with too difficult a task might not finish the experiment). For these reasons, subjects were asked to learn each list to the criterion of one perfect recitation of each pair. This measure proved reliable and informative in the work of Carroll and Burke (1965). Experimenters were provided with special scoring sheets that made it possible to see at a glance which pairs had not yet been correctly anticipated once, and made it possible for experimenters to stop the experiment as soon as the last pair had been anticipated correctly.

Ability Test. Part V of the MLAT (Carroll and Sapon, 1958) is a paired-associate task in which the student is asked to memorize as many pairs as possible out of a list of 24 pairs given to him for a brief time. The subject is then given a multiple-choice test and asked to choose the correct response to each stimulus from several alternatives. Pairs have pseudo-Kurdish words as stimuli and English words as responses.

Assignment of Subjects to Groups. Practical considerations of scheduling prevented assignment of subjects to groups by any formal scheme for randomization. Subjects names were taken as they answered an advertisement. Groups were scheduled for the several experimenters each week and then subjects were called until schedules were filled. The 10 subjects for any one group were not run in the same week; groups were spread over weeks (several groups were run concurrently). The 10 subjects in any one group were not all scheduled at the same hour of the day, though an individual subject's sessions were held at the same hour across the week.

CHAPTER III

Results and Discussion

Some of the results of the experiment followed from earlier work reviewed in Chapter I. Others were unexpected. The analyses presented below are organized according to the various variables we have dealt with. The picture of learning that they form is not complete and is, in some ways, less than clear. But many of the findings are new and, we hope valid, and should be of interest both theoretically and practically.

List Length

The majority of the experimental questions we have asked center around list length as a variable. Most of these questions are concerned with learning curves, rather than individual scores for a given day, since optimization of learning depends on information about such curves. In order to answer such questions, we need, first of all, to know what the shapes of these curves are, so that we may apply the proper methods of analysis. The original learning curves for each group may be seen in the graphs in Appendix B, and our first experimental "result" might well be the statement that either these effects are strong and remarkably consistent, or else we were extremely lucky. The curves for the various groups clearly belong to the same family of learning curves; our subjects did not generate "random numbers." Considering the small number of subjects per group (10), we are fortunate. The curves shown in the graphs in Appendix B appear to be exponential curves, and so the logarithmic transformation was applied. The results may be seen in the graph in Appendix C. All of these graphs (in both Appendix B and Appendix C) are based on a measure of learning time arrived at by the following formula:

(Trials - 1) x List Length x 4 seconds.

Multiplying list as length by four seconds gives the amount of time needed for one trial at that list length. This figure has been multiplied by the number of trials the subject used to reach criterion, less one. It was felt that the criterion trial ought not to be included, since a subject could not do better than a single trial at relearning. That is, even if he knew the list perfectly when he entered the room for the experiment, he still had to take one trial in order to prove that he knew it. If we are considering his learning times over days, this extra trial adds more to his relearning scores on the last day he sees a list than it does to his first learning trial for that list; the criterion trial might well mask the true ratio of relearning times to learning times over days for a list. The decision is, in a sense, arbitrary, but it is consistent, and results of comparisons among groups ought not to be affected by it.

The graphs in Appendix C, which show logarithmic learning times ($\log_{e}$), would seem to indicate that the learning curves for individual lists are exponential in nature, and that the logarithmic transformation renders them fairly linear. If this is true, then ordinary parametric statistical analyses can be used to answer the questions we have asked. If prediction is the aim of the experiment, we will certainly find out whether the data behaves well enough to be treated in this way. However, we do have to take into consideration, first of all, the fact that the data are correlated; we wish to deal with learning curves for each list and group, and the scores for any given list over days were all generated by the same subjects. (For example, the 42-pair control group of ten subjects provided the scores for days one, two, three, and four, as well as the final relearning scores for day five.) Analysis of these curves, or comparisons among curves from different groups requires a form of "Trend analysis" (Edwards, 1962). Further, we have indicated that ability scores will be used to adjust the learning curves of the individual subjects--necessary because the subjects selected themselves by answering an advertisement, and we have no other way of equating the groups. A trend analysis with covariance adjustment was required, and can be done using a method provided by Dempster (1965). The method is not found in the standard texts on statistics, and a brief description of the technique seems in order.

The method to be described answers questions of the form: "Are learning curves A, B, C, . . . N significantly different in means, slopes, curvature, or double-curvature (an "S" curve, or the quadratic element)? That is, it answers these questions for any curve composed of four data points. If there are only three data points (in our case, three learning and relearning days), then we cannot ask about the last term, the "S" curve. With only two points or learning days, we can only ask about means and slopes; the lists learned on the last or fourth day only can be treated by ordinary covariance methods since there is no problem about "correlated" data when each group provides only one score.

If we confine the discussion to the first list presented to students, List One, then we have four data points (the four time scores, expressed in logarithms, for learning and relearning on Days One through Four). As we have noted, these scores are correlated. If we label them T1, T2, T3, and T4 (where the subscripts designate the day of learning or relearning), we can transform them as:

$M_1 = T_1 + T_2 + T_3 + T_4$ A measure of the overall mean of the curve

$M_2 = -3T_1 - T_2 + T_3 + 3T_4$ A measure of the overall slope of the curve

$M_3 = T_1 - T_2 - T_3 + T_4$ A measure of curvature

$M_4 = -T_1 + 3T_2 - 3T_3 + T_4$ A measure of double curvature or "S" curvature

These are, of course, equations fitting a fourth-order polynomial (Fisher and Yates, 1953). A set of M Scores was generated for each student in each group, and these M Scores were used in the covariance analyses. Suppose, for example, that we wished to compare the learning curves for the lists learned by the three control groups. We have a set of M Scores for each student in the 42-pair group, a set for the students in the 26-pair group, and one for the students in the 16-pair group. The M Scores were subjected to a series of covariance analyses, using the COVAR program from Cooley and Lohnes (1962), as adapted by Jones (1964). Two measures of ability were used as covariance adjusters: The score on Part V of the MLAT and the number of correct responses made during five trials on the practice list on the first day. The analyses were carried out in the following sequence:

Table 7

Sequence of Covariance Analyses Using M Scores

	Adjusters	Dependent Variable	
1.	P-list, MLAT	M_1	(measure of means)
2.	P-list, MLAT, M_1	M_2	(measure of slope)
3.	P-list, MLAT, M_1, M_2	M_3	(measure of curvature)
4.	P-list, MLAT, M_1, M_2	M_4	(measure of "S" curvature)

As Table 7 indicates, each comparison required four covariance analyses. Analysis No. 1 compared the three control groups on the basis of their means, No. 2 compared them on the basis of their slopes after the effect due to the means was removed, No. 3 compared the groups to see if they differed in curvature after effects due to means and slopes had been removed, and No. 4 compared the groups to show differences in "S" curvature after the preceding three effects had been removed. The transformation of time scores into M Scores removes the objection that the scores are correlated with each other, since the M Scores are not. The method of covariance analysis outlined in Table 7 can be applied to situations in which only three data points are present instead of four by using the correct third-order transformations in estimating the M Scores. For situations in which only two data points (in our case, two learning and relearning sessions, as for the third day's lists) are present, only a comparison of slopes is possible. This can be done by obtaining the difference between the two points for each student and applying one covariance analysis instead of three or four, since the difference is equal to the slope when only two points are available. Theoretically, there are no limits on the method in the other direction; five or more points could be fitted to M Scores and used. It would probably prove both more practical and more interesting to use other techniques if the number of points was very large, however.

Armed with this technique, we analyzed the scores for all the first-learned lists (List One) according to the questions we had asked about them, which covered most of the questions asked in Experiments I, II, and III.

Experiment I. The groups involved in this set of analyses were the control groups at the three different list lengths. These groups received only List One, and their scores were expected to reflect learning and relearning times when no interference is present. We asked:

1. What kind of learning curves are obtained when subjects learn and relearn a single list over four days?

As we have noted, the curves are exponential in form, and are linear when transformed logarithmically-at least as linear as any psychologist could desire. Certainly they appeared to behave well enough to be useful for prediction purposes and, as will be seen, they were.

2. Do learning curves for the three control groups differ as to their means, their slopes, their curvature, and their double-curvature?

The answer is that they do differ. Table 8 gives the F-ratios for the four analyses (for means, slopes, curvature, and double curvature for these groups. The F-ratio required for significance at the .05 level are also given in Table 8.

Table 8

F-ratios for Control Group Covariance Analyses

	Means(M_1)	Slopes(M_2)	Curvature(M_3)	Double-Curvature(M_4)
Obtained F	8.73 *	4.81 *	10.39 *	1.53
Required F at .05	3.38	3.40	3.42	3.44
DF	2,25	2,24	2,23	2,22

* Significant at or beyond the .05 level

The means, both adjusted and unadjusted for all three groups in each of the above analyses are given below in Table 9.

Table 9

Adjusted and Unadjusted Group Means for Analyses One Through Four in Table Eight Plus Coefficients for Adjustment

	Group			
Dependent Variable	16-00	26-00	42-00	Coeff. for Adj.
M_1 Unadj.	15.47	19.06	22.39	P-list = -.022
M_1 Adj.	15.73	19.10	22.09	MLAT = 0.03
M_2 Unadj.	-17.63	-14.89	-14.01	P-list = 1.78
M_2 Adj.	-12.34	-15.07	-19.12	MLAT = 0.87
				M_1 = 0.05

Table 9 (continued)

Adjusted and Unadjusted Group Means for Analyses One Through Four in Table Eight Plus Coefficients for Adjustment

Dependent Variable	Group 16-00	 26-00	 42-00	Coeff. for Adj.
M_3 Unadj.	-1.39	1.07	0.61	P-list = 0.73
M_3 Adj.	-3.46	0.62	3.12	MLAT =-1.09
				M_1 = -0.23
				M_2 = -0.10
M_4 Unadj.	-0.87	-0.21	-1.22	P-list = 1.96
M_4 Adj.	0.79	-2.23	-0.86	MLAT = -0.30
				M_1 = -0.08
				M_2 = 0.18
				M_3 = -0.28

As the F-ratios in Table 8 indicate, there are differences among the groups in their means, which is not surprising. The difference in list length would certainly be expected to lead to a difference in the means. The slopes are also different. When adjusted for the effects of the P-list, the MLAT, and M_1, a measure of the means, the direction of the three groups is reversed (see M_2 adjusted and unadjusted in Table 9). The adjusted slopes are sharpest for the

42-00 group and least so for the 16-00 group. We asked, as one of our questions under Experiment I whether the slopes or learning curves for these groups would differ, and evidently, they do. The curvature also differs, according to the analysis in Table 8. This was not expected, and is not particularly good news. This measure of curvature is difficult to interpret. It represents what is left in the way of variance when the effects due to the means and slopes are removed, but it does not describe the curvature for the various groups. Further, the means for the analysis (Table 9) do not fit well with the shape of the learning curves for the control groups (See Appendix C). This curvature figure proved elusive in later comparisons of experimental groups with control groups and with each other; it appeared significant infrequently and seemingly without pattern. Whatever the figure means, it would, perhaps, be more profitable to discuss it in the light of the rest of the analyses of covariance.

Experiment II. The groups for these analyses are those in which list length was the same for all four days (groups 16-16, 26-26, and 42-42). These groups represent the largest amount of interference given in the entire study. We asked:

1. Are there differences in the learning curves of the initial list for these three groups?

Table 10 presents the F-ratios for the four analyses of covariance utilizing the M Scores for List One learning times in groups 16-16, 26-26, and 42-42.

Table 10

F-ratios for Covariance Analyses: Experiment II

	Means (M_1)	Slopes (M_2)	Curvature (M_3)	Double-Curvature (M_4)
Obtained F	32.56*	9.49*	2.19	1.36
Required F at .05	3.38	3.40	3.42	3,44
DF	2,25	2,24	2,23	2,22

* Significant at or beyond the .05 level.

Table 11

Adjusted and Unadjusted Group Means for Analyses One through Four in Table Ten Plus Coefficients for Adjustment

	Group			
Dependent Variable	16-16	26-26	42-42	Coeff. for Adj.
M_1 Unadj.	18.92	23.51	25.72	P-list = -0.11
Adj.	18.92	23.19	26.03	MLAT = -0.15
M_2 Unadj.	-10.83	-9.44	-8.29	P-list = 1.87
Adj.	- 3.17	-10.53	-14.85	MLAT = 0.38
				M_1 = 0.02

Table 11 (continued)

Adjusted and Unadjusted Group Means for Analyses One through Four in Table Ten Plus Coefficients for Adjustment

Dependent Variable	Group 16-16	Group 26-26	Group 42-42	Coeff. for Adj.
M_3 Unadj.	0.02	0.14	1.03	P-list = 0.23
Adj.	0.52	0.01	0.66	MLAT = 0.08
				M_1 = 0.06
				M_2 = 0.01
M_4 Unadj.	-1.22	-1.39	-0.02	P-list = 0.34
Adj.	0.33	-1.47	-1.50	MLAT = 0.18
				M_1 = 0.30
				M_2 = 0.06
				M_3 = 0.03

As shown in Table 10, the learning curves for these three groups on the first list learned, are different on their means, again not remarkable, and different in slope. They are not different in curvature or double-curvature. It would also appear, both in these analyses and in the earlier ones, that the two measures of ability we have used (Part V of the MLAT and the first day's score on five trials of the P-list) are not accounting for much of what is happening here. Their correlations with the various M Scores are not high; their contribution to the adjustments in the case of the analysis of M_1 are small. So far as the original question is concerned, we can say that these groups do differ in means and in slope. The next question is:

2. How do the first lists (List One) in these three groups (16-16, 26-26, and 42-42) differ from their respective control groups? If the control groups are presumed to be free of interference effects due to interpolated material, then a comparison of the first list here with the control group list at any given length would give some estimate of the amount of interference caused by the other three lists learned in these experimental groups.

Table 12 gives the results of the analyses of covariance for group 16-16 and control group 16-00.

Table 12

F-ratios for Covariance Analyses of List One in Groups 16-00 and 16-16

	Means (M_1)	Slopes (M_2)	Curvature (M_3)	Double Curvature (M_4)
Obtained F	9.25*	2.30	0.0042	0.09
Required F at .05	4.49	4.54	4.60	4.67
DF	1,16	1,15	1,14	1,13

* Significant at or beyond the .05 level

Table 13

Adjusted and Unadjusted Group Means for Analyses One Through Four in Table Twelve Plus Coefficients for Adjustment

Dependent Variable		Group		
		16-00	16-16	Coeff. for Adj.
M_1	Unadj.	15.47	18.92	P-list = -0.01
	Adj.	15.24	19.15	MLAT = -0.28
M_2	Unadj.	-17.63	-10.83	P-list = 1.53
	Adj.	-15.71	-12.75	MLAT = 0.50
				M_1 = -0.07
M_3	Unadj.	-1.39	0.02	P-list = 0.55
	Adj.	-0.67	-0.70	MLAT = - 0.75
				M_1 = - 0.15
				M_2 = - 0.07
M_4	Unadj.	-0.87	-1.22	P-list = 1.39
	Adj.	-0.84	-1.26	MLAT = -0.07
				M_1 = -0.28
				M_2 = 0.10
				M_3 = -0.13

These two groups (16-00 and 16-16) differ only in their means on the first list. This difference may be due, in part, to a difference in ability; the two measures of ability used here may not have accounted entirely for variance due to ability. The generalized test for homogeneity of variances (the H1 test of Cooley and Lohnes, 1962) yields an F-ratio of .58, at 1 and 978 df, which is not significant. There is not, then, a clear reason why these groups differ in their means. They do not differ in slope, curvature, or double-curvature. Evidently, aside from a different intercept, learning curves are alike for these two groups.

The next two groups to be compared are the 26-00 group and the 26-26 group. Table 14 gives the F-ratios for the four analyses comparing the first lists in these two groups.

Table 14

F-ratios for Covariance Analyses of List One in Groups 26-00 and 26-26

	Means (M_1)	Slopes (M_2)	Curvature (M_3)	Double-Curvature (M_4)
Obtained F	12.94*	1.45	0.006	2.30
Required F at .05	4.49	4.54	4.60	4.67
DF	1,16	1,15	1,14	1,13

* Significant at or beyond the .05 level

Table 15

Adjusted and Unadjusted Group Means for Analyses One Through Four in Table Fourteen Plus Coefficients for Adjustment

		Group		
Dependent Variable		26-00	26-26	Coeff. for Adj.
M_1	Unadj.	19.06	23.51	P-list = -0.33
	Adj.	19.22	23.35	MLAT = 0.01
M_2	Unadj.	-14.89	-9.44	P-list = 1.88
	Adj.	-10.92	-13.41	MLAT = 0.38
				M_1 = 0.11
M_3	Unadj.	1.07	0.14	P-list = 0.54
	Adj.	0.67	0.54	MLAT = -0.65
				M_1 = -0.32
				M_2 = -0.04
M_4	Unadj.	-0.21	-1.39	P-list = 2.35
	Adj.	-2.26	0.65	MLAT = -0.37
				M_1 = 0.05
				M_2 = 0.15
				M_3 = -0.04

Again the means differ significantly but the slopes, curvature, and double-curvature do not.

The third and last comparison to be made here is between control group 42-00 and group 42-42, on their first lists (the only list seen by the control group). Table 16 gives the F-ratios for all the M scores.

Table 16

F-ratios for Covariance Analyses of List One in Groups 42-00 and 42-42

	Means($_1$)	Slopes(M_2)	Curvature(M_3)	Double-Curvature(M_4)
Obtained F	7.98*	2.21	0.03	0.39
Required F at .05	4.49	4.54	4.60	4.67
DF	1,16	1,15	1,14	1,13

*** Significant at or beyond the .05 level.**

Table 17

Adjusted and Unadjusted Group Means for Analyses One Through Four in Table Sixteen Plus Coefficients for Adjustment

Dependent Variable		Group 42-00	 42-42	Coeff. for Adj.
M_1	Unadj.	22.39	25.72	P-list =-0.02
	Adj.	22.36	25.75	MLAT = -0.03

Table 17 (continued)

Adjusted and Unadjusted Group Means for Analyses One Through Four in Table Sixteen Plus Coefficients for Adjustment

Dependent Variable		Group 42-00	42-42	Coeff. for Adj.
M_2	Unadj.	-14.01	- 8.29	P-list = 1.89
	Adj.	- 9.36	-12.94	MLAT = 0.82
				M_1 = 0.01
M_3	Unadj.	0.61	1.03	P-list = 0.47
	Adj.	0.72	0.92	MLAT = -0.70
				M_1 = -0.04
				M_2 = -0.02
M_4	Unadj.	-1.22	-0.02	P-list = 1.90
	Adj.	-1.16	-0.08	MLAT = -0.37
				M_1 = 0.14
				M_2 = 0.34
				M_3 = -0.31

Again, the groups differ on their means, but not on slopes, curvature, or double-curvature. We have compared the learning and relearning of the first list (the list given on Day One and every day thereafter) in the control groups, which received no other lists, and in the three groups which received lists of a length equal to their first lists on all succeeding days. The pattern is remarkably consistent; the slopes of the learning curves, their curvature and double-curvature, are not different (if, in fact, there is any curvature, etc.). The learning of 16, 26, or 42-pair lists is apparently affected by large amounts of interference only in terms of the intercept of the learning curve, or line. It may be, as we have noted, that the two measures of ability used here were not sufficiently related to the learning task to control efficiently the effects of a "brighter" group versus a "slower" group; in all three comparisons, the non-control groups have higher M_1 Scores. (M_1 is a simple addition of times required to learn on each of the four days, for List One). Scores on the two ability tests, which will be discussed in more detail later, indicate no interesting differences between subjects in the groups here. There is, therefore, no reason to believe that the difference observed between means or M_1 Scores in each of these three comparisons was due to brighter or slower subjects, without further evidence.

In the three experimental groups discussed above (groups 16-16, 26-26, and 42-42), subjects received a total of four lists over the four days of learning. In each group, these four lists were of the same length. We asked:

3. Do the learning curves for the four different lists differ within a given group? Are there any interference effects in later lists that were not present in the first list learned? If such effects appear, they would be evidence that interference builds up cumulatively, at least when lists of equal length are used.

In order to answer this question, a series of difference scores was calculated. It is not feasible to compare whole learning curves at once, since such curves are doubly correlated--the same subject produced the scores for all days and all lists. The diagram in Figure 2 indicates the comparisons to be made for any one of the three groups under discussion.

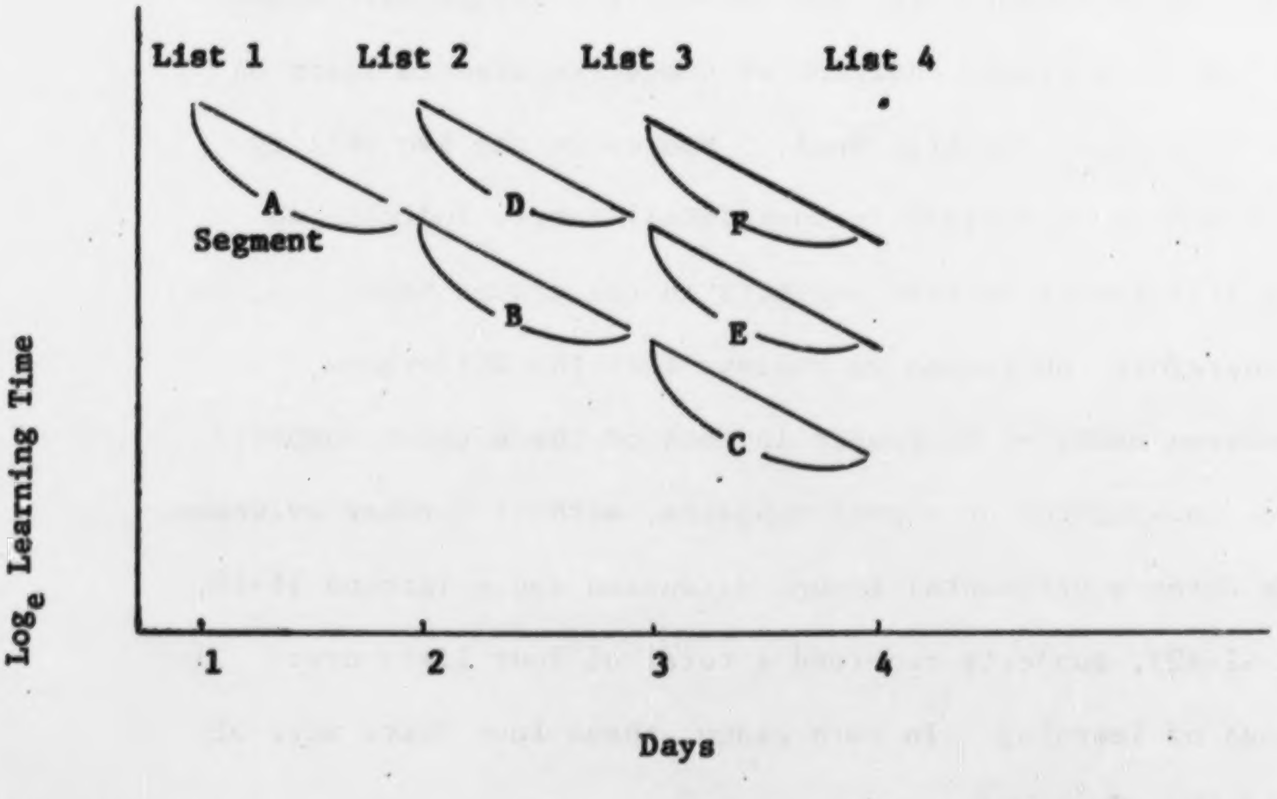


Figure 2. Diagram of lists for experimental groups indicating difference scores to be compared.

The difference between any two points on a given learning curve in the diagram in Figure 2 can be considered a measure of the slope of that segment of the line, since the line is drawn between two points and only two points. If we wish to compare subjects' difference scores for Day 1--Day 2 on List 1 (Segment A) with their difference scores for the first two sessions on List Two (Segment D), for example, we can do so by means of a simple t-test. The difference scores are still correlated; the same subjects produced both the scores being compared. But the problem of "double" correlation, which makes dealing with these curves in their entirety a problem, has been removed, and the t-test for correlated samples may be applied to the mean slopes or mean difference scores for any two segments in the diagram above, for each of the three experimental groups under discussion (16-16, 26-26, and 42-42).

The various segments of Group 16-16 were analyzed. Table 18 gives the results, utilizing the lettering of the segments in the diagram in Figure Two as a guide.

Table 18

Results of t-tests for Differences Among Slopes of Learning Curves Over Days in Group 16-16

Segments Compared	t value	ndf	p
A with D	1.95	9	ns
A with F	1.74	9	ns
D with F	.81	9	ns
B with E	.28	9	ns

For 9 degrees of freedom, the critical value of t (at the .05 level) is 2.262

The analyses above would seem to indicate that learning was not slower on second, third, and fourth lists, at least not on the segments that were compared. Table 19 presents the same analysis for Group 26-26.

Table 19

Results of t-tests for Differences Among Slopes of Learning Curves Over Days in Group 26-26.

Segments Compared	t value	ndf	p
A with D	.55	9	ns
A with F	.46	9	ns
D with F	.84	9	ns
B with E	.89	9	ns

For 9 degrees of freedom, the critical value of t (at the .05 level) is 2.262.

Again, there are no differences in the slopes of the segments compared. And finally, Table 20 presents the same analyses for Group 42-42.

Table 20

Results of t-tests for Differences Among Slopes of Learning Curves Over Days in Group 42-42

Segments Compared	t value	ndf	p
A with D	.85	9	ns
A with F	3.15	9	.01
D with F	2.34	9	.05
B with E	.91	9	ns

For 9 degrees of freedom, the critical value of t (at the .05 level) is 2.262.

This last analysis, or series of analyses presents some curious problems; Segments A and D are not significantly different from each other, yet they are both different from Segment F. (See Appendix C for plots of the segments of Group 42-42 and all other groups.) The differences or slopes of these three segments (A, D, and F) decrease across days: A = 1.35, D = 1.27, and F = 1.05. Evidently, learning was more difficult for this group on the last list, since the slope of learning from the first exposure to the third list to the second exposure is less steep than the slopes for earlier lists are. The steeper the slope, the faster the learning was, or rather, the less time was needed for relearning. This peculiar list (represented in its entirety by Segment F) comes at the outer limits of the experiment. It is one of the last lists presented to the group which had the most material to learn in all. Yet it is certainly atypical of the learning curves described in Tables 18, 19, and 20. Given that we used only ten subjects in a group, some irregularities in the data were to be expected. But it is not at all clear why this list is "out of line" and there does not seem to be any way to assess its meaning from the data available. Bearing in mind that Segment F in Group 42-42 may indicate the beginnings of interference effects, it is still interesting to hazard the guess that given list lengths are learned according to some ratio that is not much changed by the presence of other material in the student's lessons. The meaning of our one odd list will have to wait for further experimentation.

We may also answer one more of our questions with some caution. We asked:

4. Do these second, third, and fourth lists differ from the control lists at the same lengths?

With the exception of one list, the answer is that they do not. We have shown that they do not differ from the first list in their respective groups; we have also shown that those first lists do not differ from the control group lists at the same length, except in their means. It follows that these second, third, and fourth lists would compare to the control group lists, segment by segment, in the same way that they compare to their own first lists, within groups.

Experiment III. The groups in this experiment were Groups 16-16, 26-16, and 42-16. These three groups received different amounts of material for the first list; all subsequent lists contained 16 pairs. There are two important questions to be asked about these groups.

1. Do learning and relearning curves differ for the first lists when the amount of "interference" in the form of second, third, and fourth lists is held constant?

Table 21 gives the F-ratios for the analyses of covariance using these three groups.

Table 21

F-ratios for Analyses of Covariance for Groups 16-16, 26-16, and 42-16

	Means(M_1)	Slopes(M_2)	Curvature(M_3)	Double-Curvature(M_4)
Obtained F	2.55	21.89*	7.26*	1.40
Required F at .05	3.38	3.40	3.42	3.44
DF	2,25	2,24	2,23	2,22

* Significant at or beyond the .05 level

Table 22

Adjusted and Unadjusted Group Means for Analyses One Through Four in Table Twenty-One Plus Coefficients for Adjustment

	Group			
Dependent Variable	16-16	26-16	42-16	Coeff. for Adj.
M_1 Unadj.	18.92	21.91	22.45	P-list = 0.09
Adj.	19.45	21.28	22.55	MLAT = -0.35
M_2 Unadj.	-10.83	- 9.97	-12.23	P-list = 2.03
Adj.	- 5.59	-11.26	-16.78	MLAT = 0.40
				M_2 = 0.11
M_3 Unadj.	0.02	0.94	0.98	P-list = 0.35
Adj.	-1.36	0.80	2.50	MLAT = -0.62
				M_1 = -0.04
				M_2 = -0.14
M_4 Unadj.	-1.22	-0.72	1.53	P-list = 1.87
Adj.	-1.13	-0.83	1.54	MLAT = -0.17
				M_1 = -0.67
				M_2 = 0.09
				M_3 = -0.17

The results of the analyses shown in Table 21 are difficult to interpret. The first lists in these three groups do not differ significantly in their means, which is surprising, since the lists are at three different lengths. Further, the groups do differ in slope and in curvature. If these results are accurate, then we would have to say that the effects of "interference lists" at 16 pairs makes learning and relearning harder for shorter initial lists. The means of the three lists (where "means" refers to the mean of the entire curve, obtained by adding all learning and relearning times for the M_1 score) are statistically equal; in other groups and other analyses, the means for shorter initial lists have been lower than those for longer lists. The adjusted slopes (see Table 22) for the three groups would indicate that relearning was relatively slower for the shorter lists. A possible explanation for these data would lie in the fact that longer lists are more practiced; students learning 42-pair lists probably received a great deal of practice on pairs they had already anticipated correctly while they were trying to "get" the last few pairs in the list. It would appear that this added practice, combined with a relatively small amount of "interference" in the form of 16-pair lists, made the longest list less difficult to learn, at least less difficult in relation to performance on 42-pair lists in other groups. The difference in curvature is most probably due to a peak in the middle of the learning curve in Group 26-16 (see Appendix C for graphs and Table 22 for adjusted M_3 means). It is the only such peak that occurs in the data, and explanations for it are not easily found. The only other occurence of a difference in curvature was found in the analyses of control group lists. The second question to be asked of these data, then, is:

2. Do the learning curves for the three initial lists differ from their respective control group lists?

We have already noted that Group 16-16 differs from Group 16-00 differ only in their means (M_1) for the first-list learning and relearning curves (Tables 12 and 13). Table 23 gives the results of analyses for Groups 26-16 and 26-00.

Table 23

F-ratios for Analyses of Covariance for Groups 26-16 and 26-00

	Means(M_1)	Slopes(M_2)	Curvature(M_3)	Double-Curvature (M_4)
Obtained F	3.22	0.26	1.05	0.96
Required F at .05	4.49	4.54	4.60	4.67
DF	1,16	1,15	1,14	1,13

* Significant at or beyond the .05 level

Table 24

Adjusted and Unadjusted Group Means for Analyses One Through Four in Tables Twenty-Three Plus Coefficients for Adjustment

Dependent Variable	Group 26-00	 26-16	Coeff. for Adj.
M_1 Unadj.	19.06	21.91	P-list = 0.20
Adj.	19.45	21.51	MLAT = -0.16
M_2 Unadj.	-14.89	-9.97	P-list = 2.01
Adj.	-11.94	-12.92	MLAT = 0.74
			M_1 = 0.06
M_3 Unadj.	1.07	0.94	P-list = 0.58
Adj.	1.56	0.44	MLAT = -0.66
			M_1 = -0.17
			M_2 = -0.15
M_4 Unadj.	-0.21	-0.72	P-list = 2.14
Adj.	-1.32	0.38	M-LAT = -0.23
			M_1 = -0.30
			M_2 = -0.05
			M_3 = -0.12

Groups 26-00 and 26-16 do not differ in any way, insofar as the first list is concerned. Table 25 gives the results of the analyses for Groups 42-00 and 42-16.

Table 25

F-ratios for Analyses of Covariance for Groups 42-00 and 42-16

	Means (M_1)	Slopes (M_2)	Curvature (M_3)	Double-Curvature (M_4)
Obtained F	0.12	1.18	1.33	0.41
Required F at .05	4.49	4.54	4.60	4.67
DF	1,16	1,15	1,14	1,13

*Significant at or beyond the .05 level

Table 26

Adjusted and Unadjusted Group Means for Analyses One Through Four in Table Twenty-Five Plus Coefficients for Adjustment

	Group		
Dependent Variable	42-00	42-16	Coeff. for Adj.
M_1 Unadj.	22.39	22.45	P-list = 0.20
Adj.	22.87	21.97	MLAT = -0.24
M_2 Unadj.	-14.01	-12.23	P-list = 1.96
Adj.	-11.97	-14.27	MLAT = 0.85
			M_1 = 0.08

Table 26 (continued)

Adjusted and Unadjusted Group Means for Analyses One Through Four in Table Twenty-Five Plus Coefficients for Adjustment

	Group		
Dependent Variable	42-00	42-16	Coeff. for Adj.
M_3 Unadj.	0.61	0.98	P-list = 0.55
Adj.	1.28	0.30	MLAT = -1.05
			M_1 = 0.09
			M_2 = -0.13
M_4 Unadj.	-1.22	1.53	P-list = 1.84
Adj.	-0.55	0.86	MLAT = -0.24
			M_1 = -0.63
			M_2 = -0.23
			M_3 = -0.33

As Table 25 indicates, Groups 42-00 and 42-16 do not differ on their relearning curves for the first list. For the three groups, Groups 16-16, 26-16, and 42-16, it may be said that the learning and relearning curves for the first list learned do not differ in any way from control group lists of equal length, with the exception of the means of the curves in the 16-00 and 16-16 pair Groups. We noted that, in the analyses of the three control groups, the M_1 means differed for the three list lengths. We also noted that the

M_1 means for these three experimental groups (Groups 16-16, 26-16, and 42-16) do not differ. Two of these three experimental groups' means do not differ from those of the control groups at equal lengths of list. Therefore, the difference between the means of the 16-00 group and the 16-16 group must be responsible for most of the difference in means for the control groups.

Experiment IV. In the light of the rather complicated picture of interference effects provided by the analyses we have presented so far, the last experiment, Experiment IV, may not provide as much information as we originally hoped it might. The groups for this set of analyses are the groups in which the first lists were of the same length (42 pairs) and the second, third, and fourth lists were varied--one group received three lists at 16 pairs, one group received the "interference lists" at 26 pairs, and one at 42 pairs in length. The 42-16 pair group and the 42-42 pair group have been studied in some of the earlier analyses. Only the 42-26 pair group had to be added, and the relative economy this provided was one of the reasons that Experiment IV was added to the study. The results of the earlier "experiments" indicate that, for a complete picture of the effects of different amounts of "interference" on first lists, we would need to add groups whose initial lists were at 16 and 26 pairs. We asked in these analyses:

1. Do the learning and relearning curves for the first

lists in these three groups differ? (All three groups received a 42-pair first list)

Table 27 presents the results of the analyses of covariance.

Table 27

F-ratios for Analyses of Covariance for Groups 42-16, 42-26, and 42-42

	Means (M_1)	Slopes (M_2)	Curvature (M_3)	Double-Curvature (M_4)
Obtained F	3.02	0.95	0.53	0.83
Required F at .05	3.38	3.40	3.42	3.44
DF	2,25	2,24	2,23	2,22

*Significant at or beyond the .05 level

Table 28

Adjusted and Unadjusted Group Means for Analyses One Through Four in Table Twenty-Seven Plus Coefficients for Adjustment

	Group			
Dependent Variable	42-16	42-26	42-42	Coeff. for Adj.
M_1 Unadj.	22.45	25.56	25.72	P-list = -0.13
Adj.	22.89	24.98	25.86	MLAT = 0.18
M_2 Unadj.	-12.23	-8.87	-8.29	P-list = 1.76
Adj.	- 9.15	-9.64	-10.60	MLAT = 0.32
				M_1 = 0.05

Table 28 (continued)

Adjusted and Unadjusted Group Means for Analyses One Through Four in Table Twenty-Seven Plus Coefficients for Adjustment

	Group			
Dependent Variable	42-16	42-26	42-42	Coeff. for Adj.
M_3 Unadj.	0.98	1.25	1.03	P-list = 0.05
Adj.	0.89	1.25	1.11	MLAT = -0.11
				M_1 = -0.04
				M_2 = 0.03
M_4 Unadj.	1.10	0.02	-0.02	P-list = 0.47
Adj.	1.38	-0.08	-0.19	MLAT = 0.00*
				M_1 = 0.00*
				M_2 = -0.00*
				M_3 = 0.12

* Computer rounding error for very small residuals yields zero.

The results in Table 27 would indicate that curves for the first list, the 42-pair list in these three groups do not differ. Evidently, for 42-pair lists, the amount of interference provided by lists of 26 pairs and lists of 42 pairs is not sufficient to make learning slower than it is when lists of 16 pairs are used for "interference."

2. Do learning and relearning curves for first lists in these three experimental groups differ from learning

times for the 42-pair control group?

We have already noted that, according to the results shown in Table 25, Groups 42-00 and 42-16 do not differ on their first lists in any way. Table indicates that Groups 42-00 and 42-42 differ only on the means for the M_1 scores. Table 29 gives the results of the analyses of covariance comparing the first lists in Groups 42-00 and 42-26.

Table 29

F-ratios for Analyses of Covariance for Groups 42-00 and 42-26

	Means (M_1)	Slopes (M_2)	Curvature (M_3)	Double-Curvature(M_4)
Obtained F	10.13*	0.004	1.13	0.29
Required F at .05	4.49	4.54	4.60	4.67
DF	1,16	1,15	1,14	1,13

*Significant at or beyond the .05 level

Table 30

Adjusted and Unadjusted Group Means for Analyses One Through Four in Table Twenty-Nine Plus Coefficients for Adjustment

	Group			
Dependent Variable	42-00	42-26	Coeff. for Adj.	
M_1 Unadj.	22.39	25.56	P-list	= -0.15
Adj.	22.37	25.58	MLAT	= 0.02

Table 30 (continued)

Adjusted and Unadjusted Group Means for Analyses One Through Four in Table Twenty-Nine Plus Coefficients for Adjustment

	Group		
Dependent Variable	42-00	42-26	Coeff. for Adj.
M_2 Unadj.	-14.01	-8.87	P-list = 1.76
Adj.	-11.35	-11.52	MLAT = 0.19
			M_1 = 0.05
M_3 Unadj.	0.61	1.25	P-list = 0.66
Adj.	0.58	1.28	MLAT = -1.12
			M_1 = -0.18
			M_2 = -0.05
M_4 Unadj.	-1.22	0.02	P-list = 2.13
Adj.	-1.09	-0.10	MLAT = -0.46
			M_1 = 0.24
			M_2 = 0.14
			M_3 = -0.20

These two groups differ only on their mean M_1 scores. The answer to Question 2 above, then, is that the three experimental groups (Groups 42-16, 42-26, and 42-42) differ from the control groups very little; learning curves for list one in the two groups with the longer interference lists differ from the 42-pair control list only in their means, and the group which had 16-pair interference lists does not differ at all from the control group.

At this point, it might be wise to summarize the findings we have presented so far.

1. Learning and relearning curves for different list lengths do differ from each other. As our control groups indicated, the overall means of the curves, the slopes, and the curvature are significantly different for 16-, 26-, and 42-pair lists when no interference is present. The method of analysis partials out differences due to means before testing slopes, and partials out both means and slopes to test curvature, by means of covariance analyses.

2. When the amount of interference (second, third, and fourth days' lists) is held constant at 16 pairs, the three different initial list lengths produce the following results:
    - A. The initial lists do not differ from their respective control groups in any way (save that the 16-00 group and the 16-16 group differ in their means).
    - B. The learning and relearning curves for the three initial list lengths do not differ in means, but do differ in slope and in curvature. This would indicate that, when 16-pair lists are used for interference, the three

      initial list lengths still behave like themselves, in that they appear similar to the control group lists at equal lengths. Thd difference noted above in control group means must be due to the 16-00 group. Table 22 gives the adjusted and unadjusted means for these comparisons, and it would appear that the adjustments made on the mean slopes and curvatures (M_2 and M_3) for the three groups (Groups 16-16, 26-26, and 42-42) were extreme, when compared to adjustments produced by these same variables in other analyses.

Thus far, it would appear that "interference" effects were minimal. It is not entirely clear why the overall means for the three initial-list learning curves xhould not be different; apparently the three list lengths have similar overall means. The difference in means found in the analyses of the three control groups appear to be due entirely to Group 16-00, and it is not surprising that one group turned up a little out of line. With only ten subjects per group, we were fortunate that the data are as regular as they are. If we recall that "mean" here denotes simply the addition of times to learn and relearn a given list over four days, then perhaps, we can offer a possible explanation for the learning behavior we have observed. Apparently, though the number of trials required to learn lists of different lengths initially is a reflection of the number of pairs in the list, the time required to relearn lists of different lengths, when interference lists were held at 16 pairs. (Of course, if original learning times are used instead of log times, total time needed to learn would be a function of list length because the slopes of the log curves are. Practical applications stemming from these data would have to be based on real-time equations.

In addition to this, Groups 42-16 and 26-16 apparently learned the initial list on the first day at approximately the same $\log_e$ number of trials (7.61 and 7.24 respectively). The graphs in Appendix C give the figures for all groups and all lists, in logarithms. Appendix B gives the same data not logged.

3. We also found that when the three initial list lengths were followed by interference lists of equal length (equal, that is, to the first list), the results were:
   - A. Learning curves for the initial lists differed from their respective control groups only in their means. Evidently, "interference" effects were confined to the intercepts for the learning curves. Their shapes remained the same as those of the control group lists of the same length.
   - B. The curves for the initial lists differed from each other in means and slopes but not in curvature. This presents a bit of a paradox. The control group lists differ from each other in means, slopes, and curvature. These groups differ from their respective control groups only in their means. Yet these groups (Groups 42-42, 26-26, and 16-16) do not differ from each other in curvature.

This paradox is followed by another. When we compared the three groups that received 42 pairs in their initial lists and whose interference lists varied, we found that:

4. A. The curves for the initial, 42-pair lists did not

differ in any way among the three groups. In other words, interference did not increase as the length of the interpolated lists increased.

B. Group 42-16 did not differ in learning curves for the 42-pair list from the 42-pair control group. Groups 42-26 and 42-42 did differ from the 42-pair control list in the means for the initial list. These initial experimental lists do not differ from each other in any way, yet two of them differ from the control group and one does not.

Some explanation for these conflicting results may lie in the fact that the analyses are performed on adjusted scores. All scores were adjusted for ability, using the first day's score on the practice list and the scores on Part V of the MLAT (Horton's adaptation) as adjusters. Comparisons of the type we have been making should be based on a stable and predictable set of adjusters. The relationship of the two ability variables to the subjects' performances in the various groups is not as stable as could be desired. As we shall see later, in the discussion of ability as a variable, the relative efficacy of these two measures of ability varied from group to group. In general, ability appears related to the amount of time needed to learn a list the first time it is seen by the subject, but does not seem to be so well-related to the amount of time needed for relearning. The exact correlation ratios varied from group to group. To a certain extent, the results of the covariance analyses may not reflect the exact learning behavior of the subjects, in that not all of the variance due to ability in each group may have been accounted for. It is impossible to tell exactly how much ability variance is not accounted for in which groups. The results presented so far are interesting and reasonably interpretable--and the method of analysis may also be of interest to experimenters. And, of course, ability variance and its problems may not be the correct explanation for the observed effects.

In any event, it seemed wisest to proceed to other methods of analysis. One of the original aims of the study was an attempt to predict students' learning behavior under the conditions used here. If we could generate a predictive equation that satisfactorily reproduces students' learning behaviors under our various conditions, then we would have some idea of what variables are involved in such behavior and how they are involved. Multiple correlation, furthermore, would enable us to assess the ability variables for the entire study; the possible objection to adjusting for ability group by group would not be present, since all groups may be analysed at once. Multiple correlation would answer most of the questions we asked, lending confirmation to the results above and answering the questions about second, third, and fourth lists that we raised in Chapter II. The variables used were:

1. Score on Part V of the MLAT (Horton's Adaptation)
2. Score (number of correct responses in five trials) on the first day's practice list
3. List length (of the given list), as 16, 26, or 42 (not transformed logarithmically).

4. Number of times a given list has been seen before (coded as 1,2,or 3.)
5. Amount of interference (total number of pairs in other lists seen before a given presentation of a given list. Does not include repetitions of interference lists, only number of pairs in them). Again, not transformed.
6. Day on which a given list was first presented (coded 1,2, or 3).

A set of these scores was calculated for every presentation of every list for each subject. A subject in an experimental group, for example, generated ten sets of scores like those above, one for the first day's list, two on the second day (for learning of the new list and relearning of List 1), three on the third day, and so forth. Subjects in the control groups generated four sets of scores apiece. Recall and relearning on the fifth day were not included in the correlation analyses; because they took place after a longer time interval and no new learning was present, these scores had to be treated differently. There were 720 sets of scores and all of these were run through the STEPWISE regression program (Jones, 1964). No options were taken; the program was allowed to choose the ordering of the variables freely. (Briefly, the program chooses the variable having the highest correlation with the criterion and produces a regression equation. It then adds the "next best" variable and produces a new equation, and so on, until all the variables have been used. The experimenter may then test for differences among the multiple correlations so produced to see if any of the variables are not contributing significantly to the equation.) The six predictors were correlated with $\log_e$ time to learn, the same measure that was used in calculating the M Scores earlier. Table 31 gives the first-order correlation matrix for the seven variables.

Table 31

First-order Correlation Matrix for Seven Variables Used in STEPWISE Regression

	(1) P-list	(2) MLAT	(3) List Length	(4) No. Times List Seen	(5) Amt.of Interf.	(6) Day List 1st Seen	(7) Time to Learn ($\log_e$)
1)	1.00	0.33	0.03	0.00	-0.10	-0.00	-0.11
2)	0.33	1.00	0.09	-0.01	0.09	0.02	-0.08
3)	0.03	0.09	1.00	0.14	0.13	-0.29	0.26
4)	0.00	-0.01	0.14	1.00	0.13	-0.48	-0.70
5)	-0.01	0.09	0.13	0.13	1.00	0.57	0.10
6)	0.00	0.02	-0.29	-0.48	-0.57	1.00	0.27
7)	-0.11	-0.08	0.26	-0.70	0.10	0.27	1.00

A correlation of .19 is required for significance at the .05 level.

The program selected variable (4), the number of times a list had been seen before as its first predictor. The Mult R was 0.6997 and the Mult R^2 was 0.4896. The B Weight for Variable (4) was -1.0634, and the intercept was 6.9438. From the matrix of partial R's, the program next selected List Length (Variable (3)). The Mult R was 0.7885, the Mult R^2 was 0.6217, and the B Weights were -1.1396 for Variable (4) and 0.0545 for Variable (3).

At this point, the program "looked" at the partials for the remaining variables and attempted to choose the "next best." The program chooses the highest number, whether it is significantly higher than any other correlation or not. In this case, there were four predictor variables left to choose from, and their partials were:

(1)	(2)	(5)	(6)
-0.1872	-0.1983	0.2374	0.0437

Positive and negative correlations are considered at their absolute value. The program, obviously, chose Variable (5) next. A t-test for the difference between the correlation for Variable (2) and the correlation for Variable (5) yielded a t-value of less than 1.00. There is no reason to believe, then, that the amount of interference (as represented by the number of pairs in lists other than the one of interest) is a "better predictor" than the MLAT. The program added these last four variables in the order shown in Table 32.

Table 32

Order of Variables Added to Regression Equation with Mult R and Mult R^2

	Mult R	Mult R^2
1. Variable (4)	.6997	.4896
2. Variable (3)	.7885	.6217
3. Variable (5)	.8019	.6430
4. Variable (6)	.8148	.6639
5. Variable (2)	.8269	.6837
6. Variable (1)	.8297	.6884

McNemar (1955, p. 279) offers a test for the significance of the difference between two multiple correlations where one correlation is based on a subset of the variables contained in the other correlation. Each of the Mult R's in Table 32 was tested against the Mult R preceding it, to see if the addition of each variable added significantly to the prediction. All of the tests were significant. However, McNemar's test is, in part, dependent on the situation, in that it utilizes the degrees of freedom represented by the number of variables and by the total number of cases in the sample. In this case, the maximum number of predictor variables is six and there were 720 cases in the analysis. McNemar's formula is:

$$F = \frac{({R^2}_1 - {R^2}_2) \;/\; (m_1 - m_2)}{(1 - {R^2}_1) \;/\; (N - m_1 - 1)}$$

m = number of variables

N = number of cases

In the analyses presented in Table 32, the difference between two correlations was divided by $(m_1 - m_2)$, a number bound to be less than six. The figure $(1 - {R^2}_1)$ was divided by $(N - m_1 - 1)$, a number close to 720, in no case less than 714. Even the smallest difference between two correlations would appear significant when the denominator of the F-ratio is as large as that, in relation to the numerator, and McNemar, Walker and Lev, Cooley and Lohnes, and Jones all caution that, though this is the best F-ratio available, it is biased in cases like the one we have presented. It is not clear, then, that all of the last four variables added to the equation offer a real increase in prediction. We have already noted that the order in which they were added is not a comment on their relative efficacy. We might have expected the ability variables, for example, to make a better showing, And, as a matter of fact, they do behave interestingly if we look at the partials that were left as the variables were added to the equations. Table 33 gives the partials for all additions after the first variable was selected (Variable (4)).

Table 33

Partial Correlations of Predictor Variables with the Criterion for STEPWISE Regression

	Partial Correlations					
Variable Added	Var (1)	Var (2)	Var (3)	Var (4)	Var (5)	Var (6)
(4)	-.1481	-.1224	.5088	.0000	.2628	-.0935
(3)	-.1872	-.1983	.0000	.0000	.2374	.0437
(5)	-.1906	-.2248	.0000	.0000	.0000	-.2422
(6)	-.1917	0.2426	.0000	.0000	.0000	.0000
(2)	-.1220	.0000	.0000	.0000	.0000	.0000

The two ability variables (Variables (1) and (2)) were apparently suppressed by List Length and the Number of Times List Seen (Variables (4) and (3)), and to some extent by the two "interference variables" too. The relationship of the two ability variables to the criterion is apparently more productive than the first-order correlations of -.11 and -.08 would imply. Table 34 gives the B Weights and the Intercept Constant for the final prediction equation, containing all six variables.

Table 34

Weights and Intercept Constant for Prediction Equation Based on Six Variables

Variables						
(1) P-List	(2) MLAT	(3) List Length	(4) No. Times List Seen	(5) Amt. of Interf.	(6) Day List 1st Seen	Intercept Constant
-0.0424	-0.0203	0.0407	-1.3966	0.0157	-0.4554	7.4303

The final R^2 was 0.6884. This represents, of course, the amount of variance we can predict in our original scores, and is a measure of how well we have done with our six predictors. We could wish for a higher Mult R^2 but this one is significant and quite respectable. The regression analysis reflects, fairly accurately, the situation described by the covariance analyses. There is some evidence of interference effects, but not a great deal. The first order correlations between the interference variables were 0.10 for "Amount of Interference" and 0.27 for "Day List First Seen." The correlation of 0.10 is not significant. The correlation of 0.27 is significant (at the .05 level), and it is lowered as soon as the two most important variables are removed, as Table 33 indicates. Some of the information contained in the day a list was first seen on, is also contained in the number of times a list has been seen before; obviously, a list presented for the first time on the fourth day has not been seen before, for example. It is too bad that we cannot get a better estimate of the real value of these lesser variables, but, on the whole, the regression equation is satisfactory. Certainly it predicts well enough to offer hope that the approach to learning used here has possibilities. We could execute a multiple regression analysis, using the two major variables and the two ability variables, leaving out the two "interference variables." Again, we would have difficulty assessing the difference between a Mult-R so obtained and the Mult R presented here with all six variables. The small set of predictors and the large number of cases renders the F-test less informative. Ultimately, however, the question of "how big is big enough?" for a Mult R is one that each experimenter has to answer for himself.

Ability

Some of the more provocative findings of the study concern the variables related to ability. We have included here several interesting phenomena that subjects' learning curves revealed. The correlations among the $\log_e$ learning times, for example, indicate that the learning times for the first exposures to the different lists are highly correlated with each other. The correlations among learning and relearning times for the same list are not so high and decrease as days, or exposures to the list, increase. Evidently, relearning times are not necessarily related to the amount of time needed to learn a given list at first presentation, nor are relearning times necessarily related to each other. Subjects are, however, consistent in the amount of time they need to learn a list initially. The correlations between first learnings of lists are, of course, higher when the two lists are of the same length than they are when the lists differ in length. Appendix H contains the tables of the correlations for all groups.

To a certain extent, the correlations between Part V of the MLAT and the time required to learn or relearn lists also reflect this relationship. By and large, the ability test is correlated highly with the amount of time needed to learn a list initially, and is not so well-correlated with relearning times. Table 35 gives the correlations for the six experimental groups.

Table 35

Correlations Between Part V of the MLAT (Horton's Adaptation) and Log_e Learning and Relearning Times for the Six Experimental Groups

		List 1	List 2	List 3	List 4
Presentation	1	-.81	-.54	-.67	-.57
	2	-.41	-.38	-.60	
	3	-.58	-.71		
(Group 16-16)	4	-.09			

Table 35 (continued)

Correlations Between Part V of the MLAT (Horton's Adaptation) and Log$_e$ Learning and Relearning Times for the Six Experimental Groups

		List 1	List 2	List 3	List 4
Presentation	1	-.75	-.80	-.74	-.78
	2	-.48	-.59	-.18	
	3	-.30	-.03		
(Group 26-26)	4	-.44			
Presentation	1	-.45	-.42	-.26	-.35
	2	*.06	*.18	* .29	
	3	-.16	-.15		
(Group 42-42)	4	* .40			
Presentation	1	-.27	-.49	-.42	-.59
	2	-.53	-.53	-.52	
	3	-.56	* .28		
(Group 26-16)	4	* .61			
Presentation	1	* .01	-.17	-.58	-.06
	2	-.33	-.19	-.09	
(Group 42-16)	3	-.09	-.10		
	4	-.17			
Presentation	1	-.76	-.71	-.48	-.49
	2	-.59	-.64	-.33	
(Group 42-26)	3	-.47	-.09		
	4	-.63			

* Positive correlation; negative one would be expected, since students scoring higher on the MLAT would be expected to learn in less time than students scoring lower on the test.

A correlation of .57 is required for significance at the .05 level.

Table 36

Correlations Between Part V of the MLAT (Horton's Adaptation) and Log_e Learning and Relearning Times for the Control Groups

		Group	
	16-00	26-00	42-00
Presentation 1	-.54	-.67	-.76
2	-.62	-.33	-.53
3	-.37	-.03	-.41
4	* .20	-.21	* .47

*

*Positive correlation; negative correlation would have been expected.

N.B. Presentations are for the four sessions on the single list given to each of these groups. List length is reflected in the group title.

A correlation of .57 is required for significance at the .05 level.

The correlations shown in Tables 35 and 36 display a general downward trend from first learning of a list to last learning of that same list. Added to the correlations among learning and relearning times, they seem to indicate that ability matters more during the initial learning of a list than it does during later relearnings of that list. Subjects were permitted to work on a list until they reached criterion, in this case, one perfect anticipation of each response. Thus, subjects who possessed some paired-associate learning ability reached criterion quickly. Slower subjects took longer to learn and received a great deal more practice on pairs already learned while trying to "get" the last few pairs. Evidently practice makes up for ability, or the lack of it, to an extent. This "finding" most certainly deserves further work. If it is true that slower students need no more relearning or practice than faster students, provided they are given enough time to master material initially, then "programming" lessons or curricula could be adjusted accordingly.

The second fact that emerges from the data in Tables 35 and 36 is the instability of the ability measure, when it is considered in its entirety, over all groups and all lists. Aside from a general downward trend over presentations of the same list, there is no discernable regularity in these data. Clearly, if we wish to use ability as a predictor under conditions such as the ones we have used, we shall have to do better. In all fairness, it should be noted again, that Harvard and Radcliffe students are a highly select and verbal population. There was a good deal of variance in some of the groups on the MLAT scores, but the overall means were high. The test may well prove more fruitful with other populations, particularly high school students. Table 37 shows the means and standard deviations on the MLAT and on the first day's session for the practice list for all groups.

Table 37

Means and Standard Deviations for All Groups on Part V of the MLAT (Horton's Adaptation) and on the First Session of the Practice List (No. of Correct Responses)

Group	P-list Mean	MLAT Mean	P-list SD	MLAT SD
16-00	15.47	21.09	3.51	3.63
26-00	19.06	20.10	2.80	2.64

Table 37 (continued)

Means and Standard Deviations for All Groups on Part V of the MLAT (Horton's Adaptation) and on the First Session of the Practice List

(No. of Correct Responses)

Group	P-list Mean	MLAT Mean	P-list SD	MLAT SD
42-00	22.39	18.80	2.83	4.10
16-16	18.92	18.50	2.70	6.96
26-26	23.51	19.09	2.59	4.67
42-42	25.72	22.50	0.93	3.10
26-16	21.91	20.40	1.42	2.55
42-16	22.45	23.50	4.66	1.58
42-26	22.56	18.80	1.22	4.87

In connection with ability, it might be noted that several groups showed a learning-to-learn effect. That is, the amount of time subjects needed to learn a new list decreased as the number of lists they had learned increased. The phenomenon can best be measured in the groups where all four lists were at the same length for given subjects. As the graphs in Appendix B and in Appendix C indicate, the total savings (from the first learning of List 1 to the first learning of List 4) were equal to approximately one-fourth of the time needed to learn List 1 initially. It would be impossible to tell, from these data, where this effect would level off. It would be worth investigating in the future, since predictions of learning time for long periods or large amounts of material would have to take learning-to-learn into account.

Errors and Omissions

All subjects' performances were scored both for errors and for omissions. Intrusions were written in and classified later, according to whether they came from:

1. Same list, stimulus given as response
2. Same list, another response given incorrectly
3. Same list, another response belonging to a stimulus with the same initial letter as the stimulus to which it was incorrectly given
4. Different list, stimulus given as a response
5. Different list, response given incorrectly
6. Subject's own invention, word given was not on any list in the experiment.

It must be noted that one of the experimenters inadvertently neglected to write in intrusions, so that data for these inquiries are based on N's of seven or eight per group. We asked:

A. Where did intrusions come from when they occurred? The complex scoring described above yielded no information, since there were almost no intrusions except responses from the same list given incorrectly, as is usual in paired-associate experiments. There were few confusions within-list, due to stimuli that began with the same letter. In all categories except the ordinary substitution of responses, the number of intrusions was extremely small, often zero (scores were means for each list on each day). If a large number of items from different lists had been "intruded" we might have some evidence that interference was operating. There was no such evidence.

B. Was there a ratio between omissions and intrusions and did it change as a function of lists or days? Again, there was nothing interesting in the subjects' behavior, except the fact that nearly all the intrusions that did occur were produced by a few students whose strategy for learning obviously included a lot of guessing. Those who started out doing a lot of guessing continued to guess on all lists and all days. It is also interesting to note that nearly all subjects differentiated between stimuli and responses quite quickly; there were virtually no stimuli given as responses either at first learning or at relearning, even in the groups which learned 42-pair lists.

Recall and Relearning

When we consider the vast amount of work that needs to be done with the variables discussed above, it seems almost presumptuous to worry about how much students remembered 72 hours after the last learning session. Certainly comparisons for the purpose of choosing the "best" group would be pointless. We might ask, however, whether we can predict a student's relearning time after 72 hours by knowing his scores on the four lists for the last learning day. Table 38 presents the correlations between learning times on the fourth day and the fifth or "testing" day for the six experimental groups.

Table 38

Correlations Between Day 4 and Day 5 Learning Times (Log$_e$) for Six Experimental Groups

Group	List 1	List 2	List 3	List 4
16-16	.56	.48	.49	* .85
26-26	.18	.25	.32	-.07
42-42	.33	*.65	.41	.37
26-16	-.07	.22	-.32	.34
42-16	-.13	.48	*.84	.35
42-26	*-.62	.13	.39	-.19

* A correlation of .57 is required at the .05 level of significance.

As Table 38 indicates, there is not much relationship between students' scores on the four lists on the last learning day and their scores on those same lists after 72 hours. For the control groups, the correlations were .67 for the 16-pair list, .25 for the 26-pair list, and .09 for the 42-pair list. The mean relearning times for the various groups are shown in the graphs in Appendix B (raw times) and in Appendix C ($\log_e$ times). There was, in general, not much loss over 72 hours.

The recall scores provided some interesting data on interference. If we consider the first trial of every relearning session as a recall test, we may obtain recall scores for all lists after their first presentation. Scores for recall are expressed as the number of correct anticipations on the first relearning trial. Since some of our groups had lists of different lengths on different days (Groups 26-16, 42-16, and 42-26), these raw recall scores were also transformed into percentages (based on the number of pairs or possible correct anticipations in the list). Table 39 presents both raw and percentage scores for recall for the six experimental groups.

Table 39

Raw and Percentage Scores for Recall on Days Two Through Five for the Six Experimental Groups

			List 1	List 2	List 3	List 4
Group 16-16	Present. (Raw)	2	6.1	3.9	3.8	3.7
		3	12.4	11.3	8.9	
		4	13.5	13.3		
		5	14.2			
	Present. (%)	2	37.9	24.4	23.7	23.1
		3	77.5	72.6	55.6	
		4	84.3	83.1		
		5	87.5			

Table 39 (continued)

Raw and Percentage Scores for Recall on Days Two Through Five for the Six Experimental Groups

			List 1	List 2	List 3	List 4
Group 26-26	Present.	2	9.0	6.1	5.7	3.7
		3	18.3	16.7	13.0	
	(Raw)	4	22.5	18.5		
		5	22.8			
	Present.	2	34.6	23.4	21.9	14.2
		3	70.3	64.2	50.0	
	(%)	4	86.5	71.7		
		5	87.7			
Group 42-42	Present.	2	19.7	20.1	15.4	16.1
		3	34.9	32.3	27.0	
	(Raw)	4	38.6	33.8		
		5	38.8			
	Present.	2	46.9	47.8	36.6	38.3
		3	83.1	76.9	64.2	
	(%)	4	91.8	80.4		
		5	92.3			
Group 26-16	Present.	2	12.1	3.0	2.3	3.0
		3	21.6	10.3	8.0	
	(Raw)	4	24.4	11.7		
		5	23.7			
	Present.	2	46.5	18.7	14.4	18.0
		3	83.0	63.2	47.8	
	(%)	4	93.8	67.1		
		5	91.1			
	Present.	2	27.6	3.0	2.7	6.0
		3	37.9	10.5	7.3	
	(Raw)	4	37.7	12.5		
		5	40.0			

Table 39 (continued)

Raw and Percentage Scores for Recall on Days Two Through Five for the Six Experimental Groups

			List 1	List 2	List 3	List 4
Group 42-16						
	Present.	2	65.4	18.7	16.8	13.7
		3	90.2	65.6	50.0	
	(%)	4	93.6	78.1		
		5	95.2			
	Present.	2	32.0	13.0	14.0	8.0
		3	40.0	23.0	21.0	
	(Raw)	4	41.0	26.0		
		5	42.0			
Group 42-26						
	Present.	2	76.2	50.0	53.8	30.8
		3	95.2	88.5	80.8	
	(%)	4	97.6	1.0000	1.000	
		5	1.000			

Table 40

Raw and Percentage Scores for Recall on Days Two Through Five for the Three Control Groups

		16-00	26-00	42-00
Present.	2	10.0	19.3	31.1
	3	15.0	24.4	39.7
(Raw)	4	15.0	25.0	40.5
	5	15.0	25.6	40.8
Present.	2	62.5	76.9	74.0
	3	93.7	93.8	95.0
(%)	4.	98.1	96.5	96.6
	5	99.3	98.4	97.1

If we compare the recall scores in the experimental groups, particularly the percentage scores, with their equivalents in the control groups, it is clear that some interference effects appeared at recall, although, as we have seen, they disappeared during relearning. That is, the relearning scores did not show so much interference; whatever lack of memory subjects displayed on the first trial for a list, they quickly recovered by the second or third trial. This temporary "interference effect" is particularly marked at the second exposure to the various lists. Statistical comparisons among percentage scores are usually difficult to do; percentages are seldom distributed normally. The differences between control-group scores and experimental scores here is pronounced enough to be worth investigating further. One possible variable is the amount of time between the different lists. We allowed the subjects only two minutes between lists. All of the figures in Table 39 are results for lists that were presented only two minutes after a new list or after relearning of another old list. It is possible that the interference effect reflected in these data is due to the very short time interval between lists. If so, then it is not particularly interesting. If such an effect were to appear even after generous intervals between lists had been used, then we would have a way of getting at interference that might be most useful.

Conclusions

How does all this help our hypothetical student, who must learn efficiently every day? The purpose of this study was the optimization of learning over time and over large amounts of material. We wished to find out what variables are involved in students' learning under these conditions and how these variables are related to each other. We would wish to consider, then, three related issues:

1. What do we know about the parameters of learning under our conditions?
2. Is the design, the method employed here, a fruitful line of attack on the problems and questions we raised?
3. Assuming that we have some useful information about learning, what should the next step be? What questions still need to be answered, and which of these should be investigated next?

We do have information about the parameters of learning under our conditions; some of it is most interesting. We noted that scores on the ability variables did not net as much information as they might have, but we also noted that scores for the initial learning of lists were highly correlated with the ability tests, while relearning scores were not so well-correlated with the tests. Given these data and the correlations among learning and relearning times within groups, it seems clear that ability is reflected in initial learning much more than in relearning. Evidently, if students of varying ability are allowed whatever amount of time they need to have for initial learning, their relearning times will be similar. It does seem feasible, then, to work with "ability" in terms of "time required to learn." It also appears that educational planning -- of programs, curricula, text, school schedules -- might be better for being more flexible as to the amount of time allowed for initial learning. If, in fact, practice of materials already presented can be alloted the same amount of time for all students, and only for initial learning need students be considered individually, then planning of materials and schedules would, perhaps, be simpler than it would be if students had to be considered individually at all points in the lessons. Such possibilities should be investigated; one of the first "follow-ups" of this study might well be a study of the relationships of various ability measures (as many as could be found) to the experimental tasks given here and a careful variation of different amounts of practice after the student has initially learned a list. Such a study would, hopefully, add to our rather meagre information about ability and its role in the optimization of learning.

This role cannot be overestimated. We have already noted that the covariance analyses of learning curves were, most probably, obscured by the rather inefficient and unstable control we had over ability. If ability has a peculiar relationship to learning and relearning times, then it may have biased the results of the multiple correlations as well. A Mult-R^2 of .69 is not high, however, for our conditions. If there is a place for ability, as a variable, in the "theoretical equation" on which this study was based, then further work should lead to a better measure of ability and better prediction. There is, in the data presented here, no evidence that ability does not enter into optimization of learning.

The role of interference in the learning and relearning of materials under our conditions is not completely clear. We noted that it was difficult to tell just how much the two interference measures ("Number of items seen in other lists" and "day the list being learned was first seen") contributed to the prediction equation. We do know that their first-order correlations with the criterion (time used to learn or relearn) were not significant, even with the large N we used. The equation is tentative; as we point out above, much more work must be done with all of the hypothesized variables.

The covariance analyses present much the same picture of the role of interference in the optimization of learning. Signs of it occur here and there, and precise location of the point in learning at which interference appears is not yet known. One suggestion is found in the results of the t-tests which were used to compare lists of the same length learned by the same group over four days (groups 16-16, 26-26, 42-42). All curves were alike, within group, except the last 42-pair list. If this is a genuine interference effect, and not due to variation among subjects or the small size of the groups, etc., then it suggests that we ought to work with longer lists over longer time spans. This particular list was the last list learned by the group with the most difficult task in the study. What is beginning to happen at this point, we cannot say now.

We also found that these three groups (16-16, 26-26, and 42-42) differed from their respective control groups only in the means for the first list learned. If we recall that differences in list length may facilitate list differentiation, then it may be that students in these groups had less help in differentiating lists than did students in groups with interference lists at lengths different from the initial list. This equality of list length may have had the effect of raising the intercept (in comparison with the control groups) without changing the shape of the learning curve that is "characteristic" of a given list length (defined by the control group learning curves). Or this difference in overall means may be a sign of interference effects due to the amount of material -- not in the sense that lists were of equal length, but in the sense that most of these subjects had more to learn than did other groups. Certainly other groups, groups with unequal list lengths, occasionally had overall means that differed from the relevant control group means, but the pattern is clearest for these three (16-16, 26-26, and 42-42).

Given the many problems and questions about the variables involved in learning that still need to be answered, it seemed pointless to speculate extensively about recall; particularly, it seemed that an attempt to include a term for recall in the tentative equation for learning might be more confusing than enlightening. It was clear from the mean recall and relearning scores (see

Appendix E and Appendix F for figures on recall) that our subjects learned satisfactorily -- perhaps even overlearned some of the materials, especially the first list given. Nearly all subjects in all groups needed only a trial or two on the fifth or testing day to relearn the first list (see Appendix B and Appendix C for raw and $\log_e$ relearning times). The data we have on recall and relearning might well serve as a first step in assessing the optimal amounts of practice needed to insure maintenance of items in the student's repertoire. We need to work on the variables involved in that learning process first, however.

The review of past paired-associate research presented in Chapter I points up the fact that much of what was covered in this study has not been done before. Some of our results are perfectly in accord with past findings. For example, the effects of interference reflected in the first trials of the various lists on all days is most likely a function of the two-minute interlist interval; it could have (and probably should have) been predicted from what is already known about the way in which associations to earlier lists recover as the time span between interference material and testing of the original list increases. Had we used a longer interval, the experimental recall scores might have compared more favorably with the control scores.

There were some indications in various past studies of interference, that lists in which both stimuli and responses were different fared better than lists with identical or similar stimuli, but were not learned so quickly as were lists with similar responses. We cannot make an exact parallel between our conditions and those used in such past studies. The facilitating effect of similar responses may not reflect an interference gradient. It may be that, as researchers have hypothesized, paired-associate learning has two phases. In the first phase, the subject "integrates" the responses; that is, he puts together the letters of an unfamiliar nonsense syllable in some way so that it becomes a word for him and not three separate letters. If the materials are words, they must still be differentiated, etc. Then, in the second phase, the subject puts together the pairs. If this explanation of paired-associate learning is correct, then a second list with responses similar to those in the first list would not require so much effort to "integrate." Unfortunately, educators are seldom able to select only similar materials to teach. Sooner or later, all material must be covered. But the relative difficulty with which second lists containing all new responses are learned may not be evidence of interference, when compared to lists with similar responses. Most of the experiments done in the area of stimulus and response similarity and interference simply compared relearning times for the relevant list (depending on whether the experiment was proactive or retroactive). One study (Bugelski and Cadwallader) did show some interference in the A-B, C-D paradigm when scores were compared with those of a control group, however. But the massive amounts of interference found in the classic A-B, A-C paradigm need not frighten educators. Most of these classic studies used lists considerably shorter than ours. Our subjects demonstrated fairly clearly that this massive interference is a function of the paradigm. Most of our subjects learned much more material than subjects are usually given, and learned it fairly easily. Their error patterns were not different from what is usually found in paired-associate studies -- some intrusions of intra list responses and very few interlist intrusions. And our subjects' patterns hold for four lists.

We might note here, that some caution should be used in judging the behavior of our subjects. Experimenters have noticed that Harvard and Radcliffe College students, even freshmen, learn verbal materials very easily, perhaps more easily than other college populations and, particularly, high school populations. Their use as subjects was unavoidable in this study, but at least a spot-check should be made, using other populations. Our subjects' learning rates may be higher than those of other groups of people, but there is no sound reason to believe that comparisons among our groups are invalid, however.

The results of this study cannot be directly compared to most of the classic interference studies, since few of those deal with the A-B, C-D paradigm and few deal with more than two lists. Many earlier studies did not even include relearning data for both lists used. At best, then, we could make some generalizations about an interference continuum. Some comparisons between our data and the data from such studies as Bugelski's and Cadwallader's, Waters' and others may be possible, and so may comparisons with studies on meaningfulness, such as those by Cieutat, Underwood and others, provided we can establish parallels of difficulty levels for materials and obtain data where the published accounts do not provide it. But, with the exception of a few comparisons, already noted, "interference" effects were not marked in our data. By and large, lists of any one given length behaved much the same under differing interference conditions; both group-by-group and day-by-day comparisons of learning and relearning times yielded more similarities than differences among lists of equal length. We have already noted the lack of a statistically significant relationship between the "interference" variables and learning time in the first-order correlation matrix of the seven variables. And for the differences that appeared in the covariance analyses there are, as we noted, several explanations, all possible. The interpretation of these results is a matter of subjective judgment (if "interpretation" means judgments about which problems to study next, which direction to choose in further research). As a judgment of the entire picture presented by the data, then, it would seem that the most interesting possibilities be in further manipulation of the ability variables and in longer list lengths. For lists of 42 pairs or fewer, and under the given conditions, interference does not seem to be a problem. Given some improvements in the predictive usefulness of ability measures and a better idea of how they work over time, the equations for 42 pairs or fewer might well be written without interference terms, without a significant loss of information. If, on the other hand, further work with longer lists brings to light an interference pattern, then we would wish to look again at the data presented here to see if the effects we noticed here and there in the comparisons are really glimmerings of an interference curve. The data we have do not permit choices among alternative explanations, nor do they permit much generalization about an hypothesized interference curve.

We asked whether the method or approach to optimization used here had value for further experimentation. Certainly our data suggest any number of possibly fruitful directions for such further work. We have already mentioned the ability variable, longer lists, and a variation of amounts of practice at different list lengths (assuming that the student is permitted to learn initially at his own speed). We could add to this list the list length variable itself. Certainly the data given here indicate that there is something to be gained in working with list length as a continuum. With only three list lengths it is difficult to specify parameters for a "list length curve" over all our other variables. But we have demonstrated that, under most conditions, the three list lengths used behave differently. The old whole-part approach does not recognize this effect. Contemporary researchers in "whole-part" have recognized the fact that there may be such a phenomenon as an "optimal block size" but their arbitrary choice of amounts of material can, at best, lead to a discovery of that optimal size by hit-or-miss methods. Supples suggests two parameters to be used in calculating block size. Such block size is a function of the probability that an item will be remembered versus the probability that it will be forgotten.

Stated this way, his parameters must include information about the number of items in a list, the student's ability, and, one would assume, the likelihood that this "block" or this "item" will be affected by whatever blocks or lists, if any, the student has already seen (this last must be included until we have proved definitely that no interference takes place under given circumstances). Suppes has not yet worked out an exact method for calculating his two probabilities, and it would seem that he would have to do something very like what we have done in order to calculate them. At least, the same variables are involved.

List length produced one interesting though unsupportable notion. There were four experimenters working on the study in all. All four administered lists to some subjects in each group. And all four noticed the difference in subjects' attitudes that corresponded with the length of the lists they were given. For what they are worth, the experimenters" subjective observations were that subjects employed quite different strategies for different list lengths. Subjects with 16-pair lists, for example, were markedly more casual and less curious about the experiment. Subjects with 42-pair lists were, almost invariably, concerned with and proud of their strategies for "hooking-up" pairs, and eager to describe them. (They were gently restrained from discussing lists until the experiment was over). Many of these subjects came by the laboratory to find out what the results of the study were and, in general, took a proprietary interest in it. Many brought in lists of the associative "tricks" or schemes they had used, voluntarily. This concern with strategies for learning was also found in the slower 26-pair list subjects. Granted that this is a subjective observation and subject to hopeful speculation on the part of the experimenters. Still, while not a "finding" it suggests that it might be most interesting to investigate subjects' strategies as a function of the amount of material they are asked to learn. Experiments in mediated verbal learning are one example of such investigation; perhaps it might be feasible to combine their techniques with our variables. Or other techniques could be developed.

The data presented here and the equations derived from these data are much too tentative to be offered to educators as practical solutions to their problems. Again, subjective judgment enters into interpretation, but, given our results, it would seem that we could suggest to educators that the approach, if not the particular parameters, is worth considering. No matter how much research is done in the laboratory, individual educators will have to work with their own materials, their own populations, etc. in order to make learning in specific situations as efficient as possible. The most we can do is identify variables and their relationships to each other and to the criterion, and we can suggest ways of working with these variables.

And to the theory, or body of knowledge about paired-associate learning, we have, perhaps, added something. As we pointed out in Chapter I, before we can explain why or how subjects do something, we must know what they did. We have been most fortunate -- the data here is regular enough to permit speculation about learning under our conditions, at least. We now know something about what subjects will do under our conditions and these conditions had not previously been tested in the particular combination we used. We have made some connections to past research and have offered some explanations, not necessarily inconsistent with past research, about why learning curves look the way they do here. And we have been able to offer several suggestions for future research problems that would extend and make more useful the results given here.

Appendix A

Materials Used in Experimental and Practice Lists

PRACTICE LIST *

Stimulus	Response
A	92
F	71
P	35
S	27
M	16
Q	63
D	54
L	48

* Two randomizations were used to prevent serial learning.

EXPERIMENTAL STIMULI

AFRAID
ARM
ANSWER
APPLE
AIR
ALLOW
ADD
ALONE

BEAUTIFUL
BUILD
BANK
BREAD
BELIEVE
BALL
BEGIN
BIRD

CALL
CITY
CHURCH
CORN
CHILD
COUNTRY
CARRY
COLOR

DEEP
DINNER
DRAW
DAY
DRINK
DIVIDE
DEAD
DRESS

EAR
EQUAL
EAST
EYE
END
ESCAPE
EARTH
EGG

FIND
FISH
FALL
FRONT
FIELD
FRESH
FAMILY
FOUR

GREEN
GAME
GOLD
GROUND
GREAT
GUIDE
GARDEN
GO

HELP
HAPPEN
HUSBAND
HORSE
HEAR
HUNDRED
HAIR
HAPPY

KISS
KEPT
KILL
KEEP
KNEE
KIND
KING
KNOW

LIE
LEARN
LAUGH
LETTER
LATE
LIGHT
LAW
LINE

MILE
MOTHER
MARK
MEASURE
MONEY
MIND
MATTER
MILK

NEXT
NOISE
NAME
NEW
NOTHING
NEST
NEED
NORTH

OUT
OBJECT
ONE
OFFER
OWN
OCEAN
OLD
ORDER

PEOPLE
POOR
PAPER
PLANT
POINT
PRACTICE
PASS
PICTURE

RIGHT
ROOM
RAIN
ROUND
RICH
REASON
READ
REMEMBER

SHIP
SOLDIER
SCHOOL
SERVE
SECOND
SILVER
SAW
SEND

TEN
TRAIN
THINK
THOUSAND
TODAY
TURN
THING
TIME

WELL
WHEAT
WISH
WORK
WARM
WINDOW
WATER
WALK

* HISTORY
RAPID
SURPRISE
CROSS
PLAY
FIRE

* SWEET
ROCK
FOOT
MASTER
HEIGHT
ART

* NARROW
FORGET
BROWN
COAT
DESIRE
POUND

* POWER
WOMAN
COVER
MOUNTAIN
LITTLE
GENERAL

* Sets of words added to each of the four forms (A, B, C, and D, respectively) to make 42 stimuli per list. For each 42-pair list, six letters of the alphabet were thus repeated three times, because of a shortage of Thorndike-Lorge A and AA words which were nouns, verbs, or adjectives.

EXPERIMENTAL RESPONSES

ANODE
AXIOM
ALACK
ARRAS
ANENT
ATTAR
ANISE
APHID

BEDEW
BIGHT
BLEAR
BASIL
BANNS
BAULK
BLUET
BOGGY

CAPON
CAIRN
CLACK
COZEN
CAROM
CHEEP
CHINE
CIVET

DOUSE
DUCAT
DAVIT
DELFT
DEIST
DOLOR
DOWDY
DACIA

ETUDE
EERIE
EGRET
EIDER
EDUCE
ELITE
EPHOD
EVICT

FIBRE
FACET
FAGOT
FLUME
FEMUR
FESTA
FETID
FITCH

GROAT
GNOME
GENIE
GRUEL
GAUZY
GAVEL
GHOUL
GORSE

HOURI
HORNY
HOLLA
HILUM
HAWSE
HENNA
HOUGH
HUTCH

IDIOM
INGOT
ILIUM
INDUE
INAPT
INDUS
INLAY
IONIA

JABOT
JUNTO
JUMBO
JOIST
JEWRY
JINGO
JULEP
JAFFA

KRONE
KRAAL
KULAK
KETCH
KAPPA
KIOSK
KABUL
KAPOK

LEAPT
LLANO
LIMBO
LEACH
LARES
LAZAR
LUPUS
LIANA

MIDGE
MERLE
MILCH
MANSE
MARDI
MAVIS
MOIRE
MOULT

NADIR
NITRE
NONCE
NATAL
NATTY
NEGUS
NILUS
NEATH

ORATE
ODIUM
OATEN
OVATE
OBESE
OMEGA
ORIEL
OZONE

PHIAL
PRAWN
PRONG
PAYER
PATEN
PROEM
PHILO
PHLOX

QUIRT
QUASH
QUEUE
QUINT

RANGY
RENAL
REFIT
REBEC
RATHE
RAVIN
REBUS
ROWEL

SOUSE
SAHIB
SWALE
SULLY
SCAPE
SCRAG
SEDUM
SPLAY

TRIPE
TAUPE
TRYST
TUNNY
THEGN
THORP
TOQUE
TRIER

UNCUT
UNBAR
UMBER
UNARM

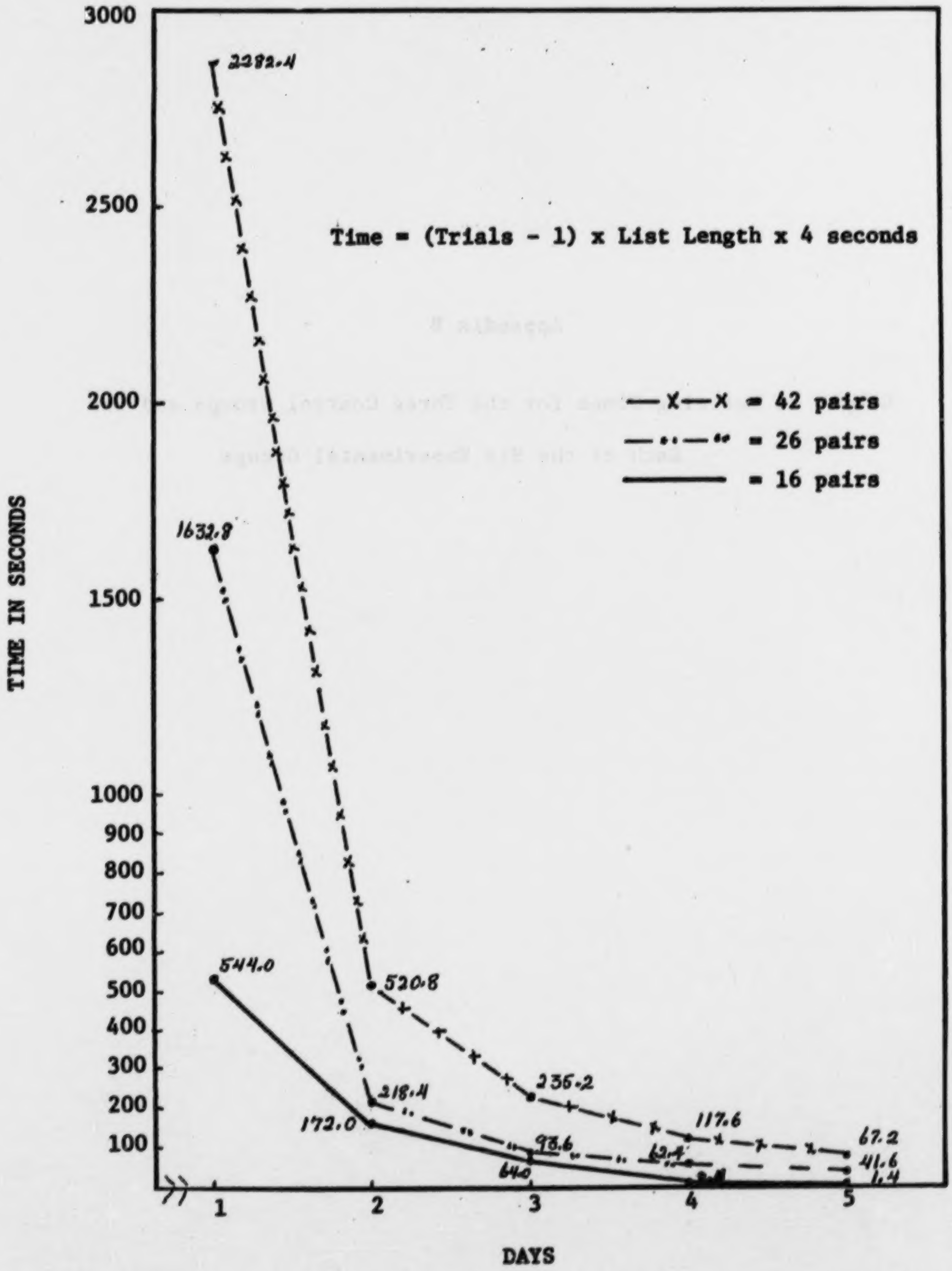
VENAL
VISOR
VYING
VETCH

WINCH
WOOER
WHELK
WELCH

Appendix B

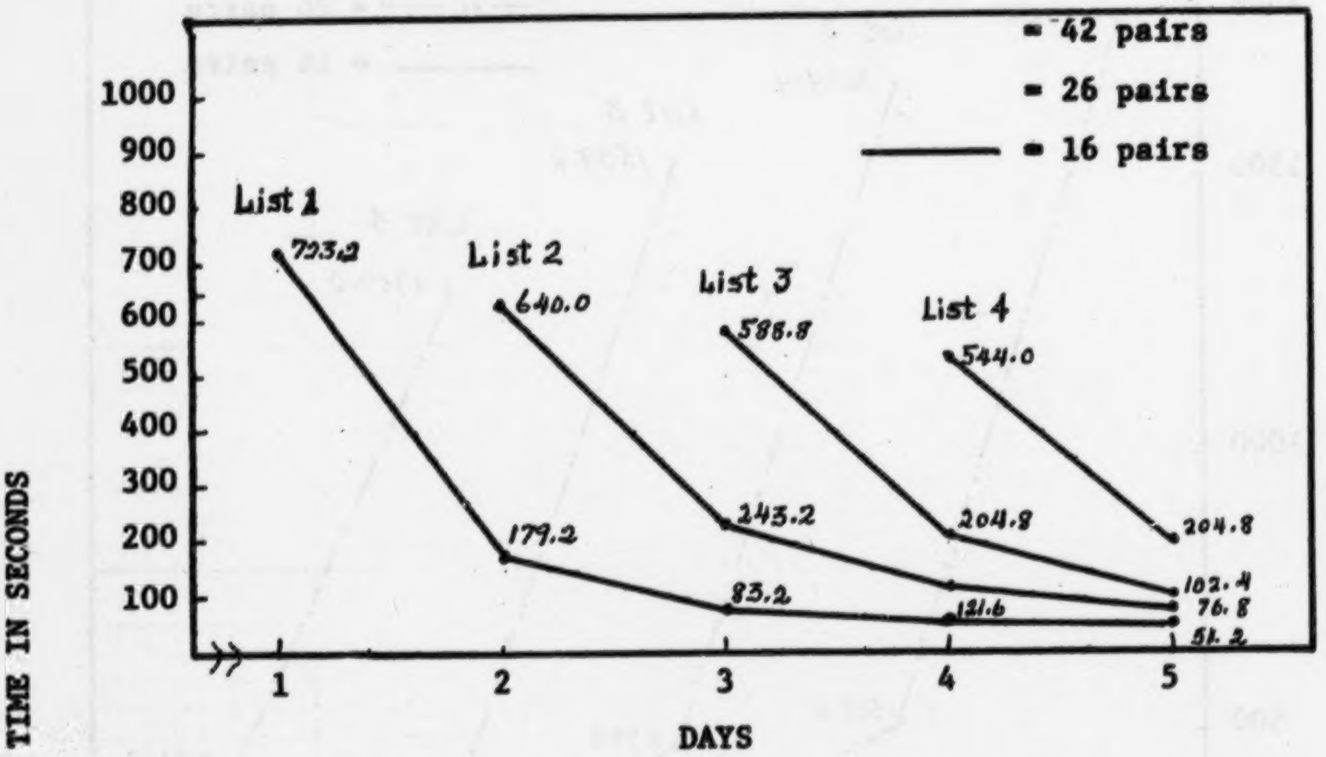
Graphs of Learning Times for the Three Control Groups and for Each of the Six Experimental Groups

MEAN TIMES TO LEARN AND RELEARN PAIRS ON DAYS 1 TO 5 : CONTROL GROUP



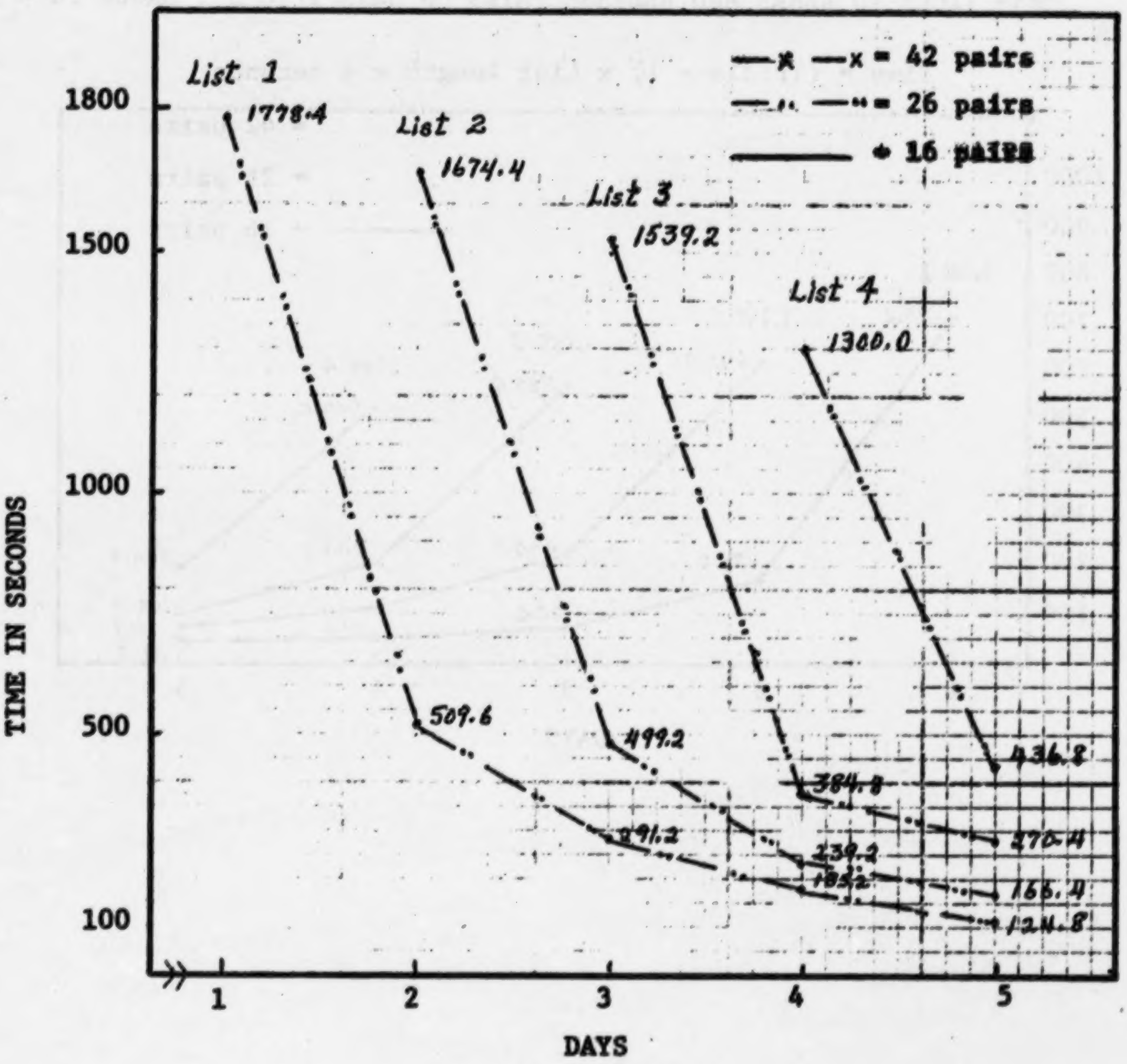
MEAN TIMES TO LEARN AND RELEARN PAIRS ON DAYS 1 TO 5 : GROUP 16 - 16

Time = (Trials - 1) x List Length x 4 seconds



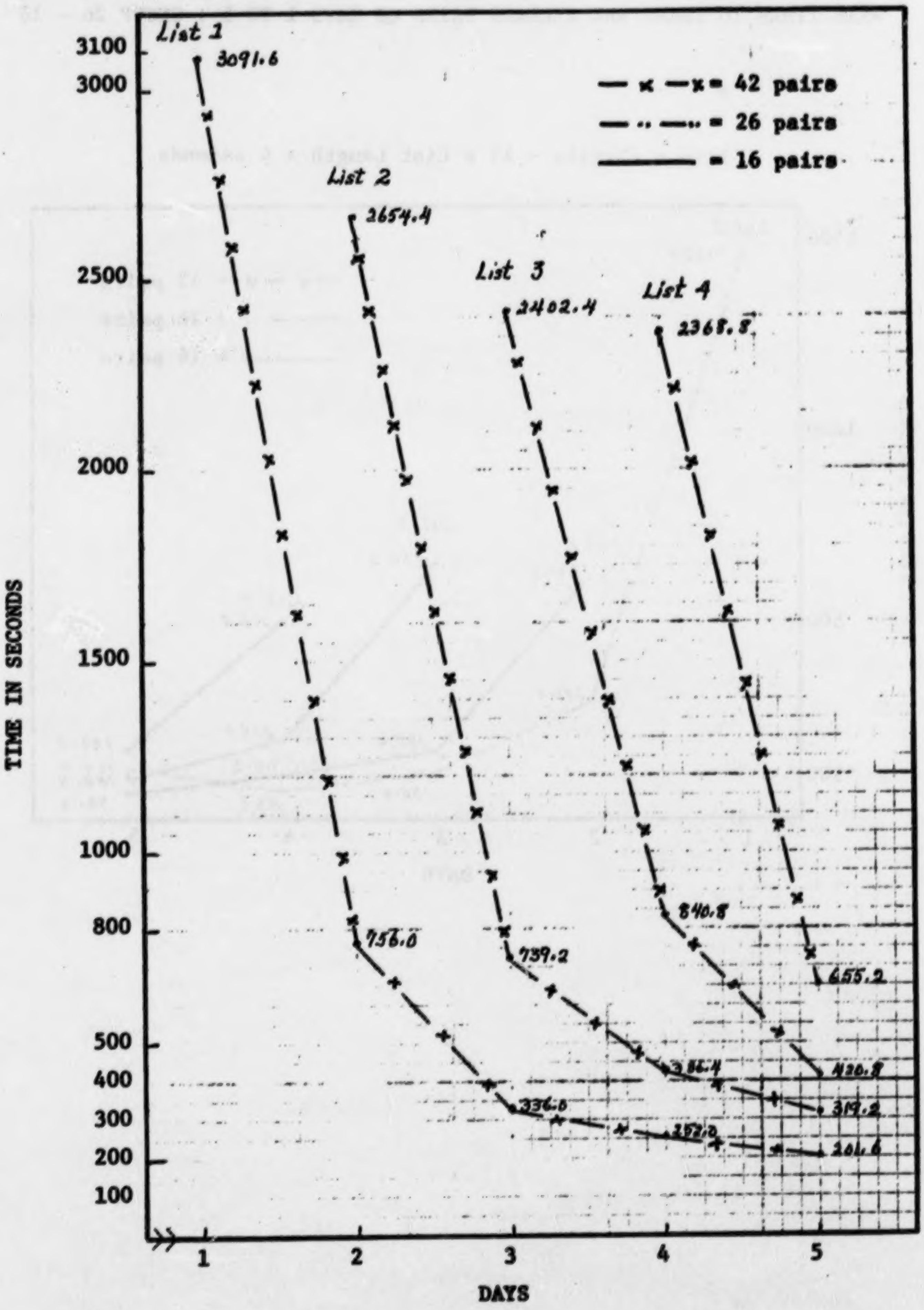
MEAN TIMES TO LEARN AND RELEARN PAIRS ON DAYS 1 TO 5 : GROUP 26 - 26

Time = (Trials - 1) x List Length x 4 seconds



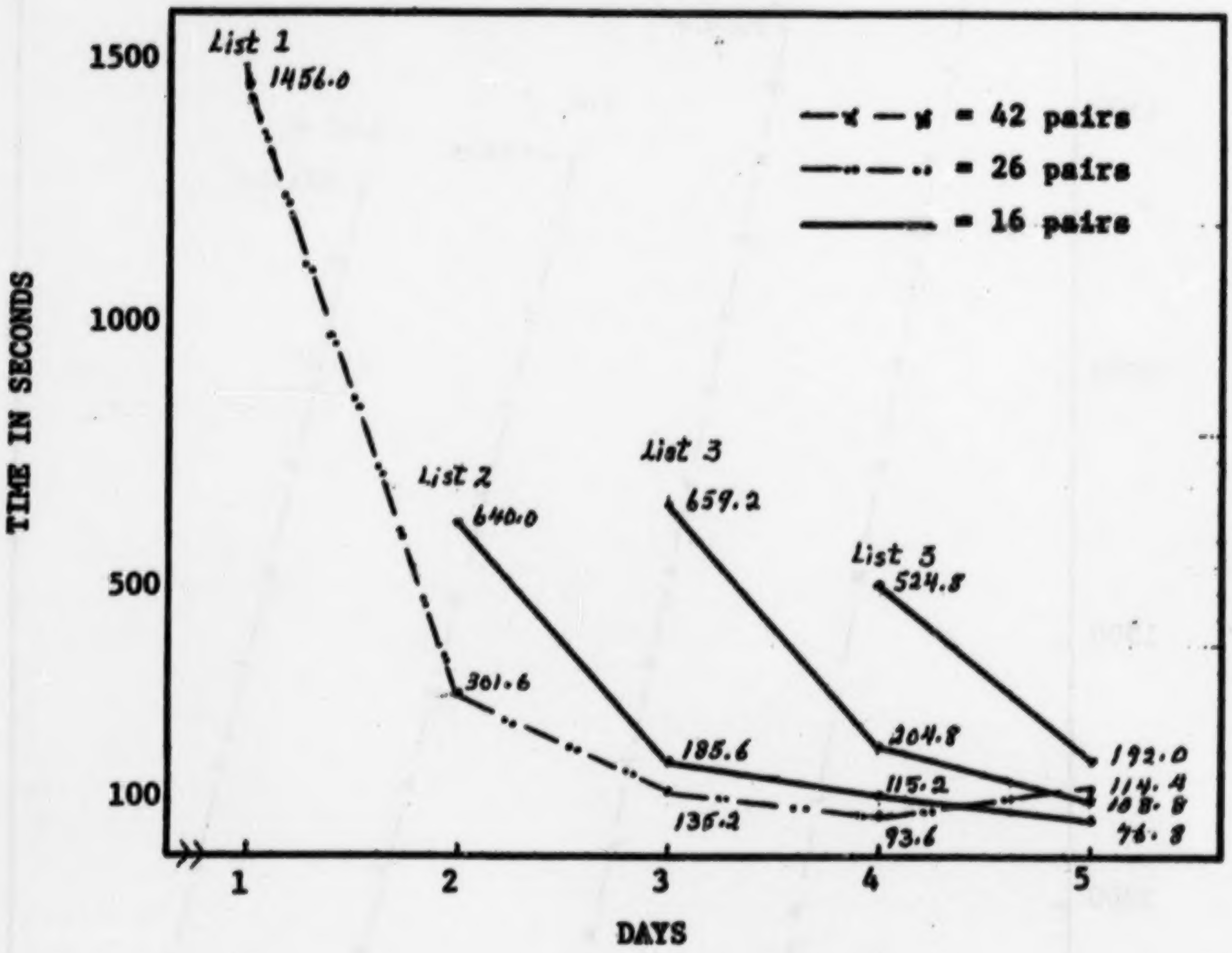
MEAN TIMES TO LEARN AND RELEARN PAIRS ON DAYS 1 TO 5 : GROUP 42 - 42

Time = (Trials - 1) x List Length x 4 seconds



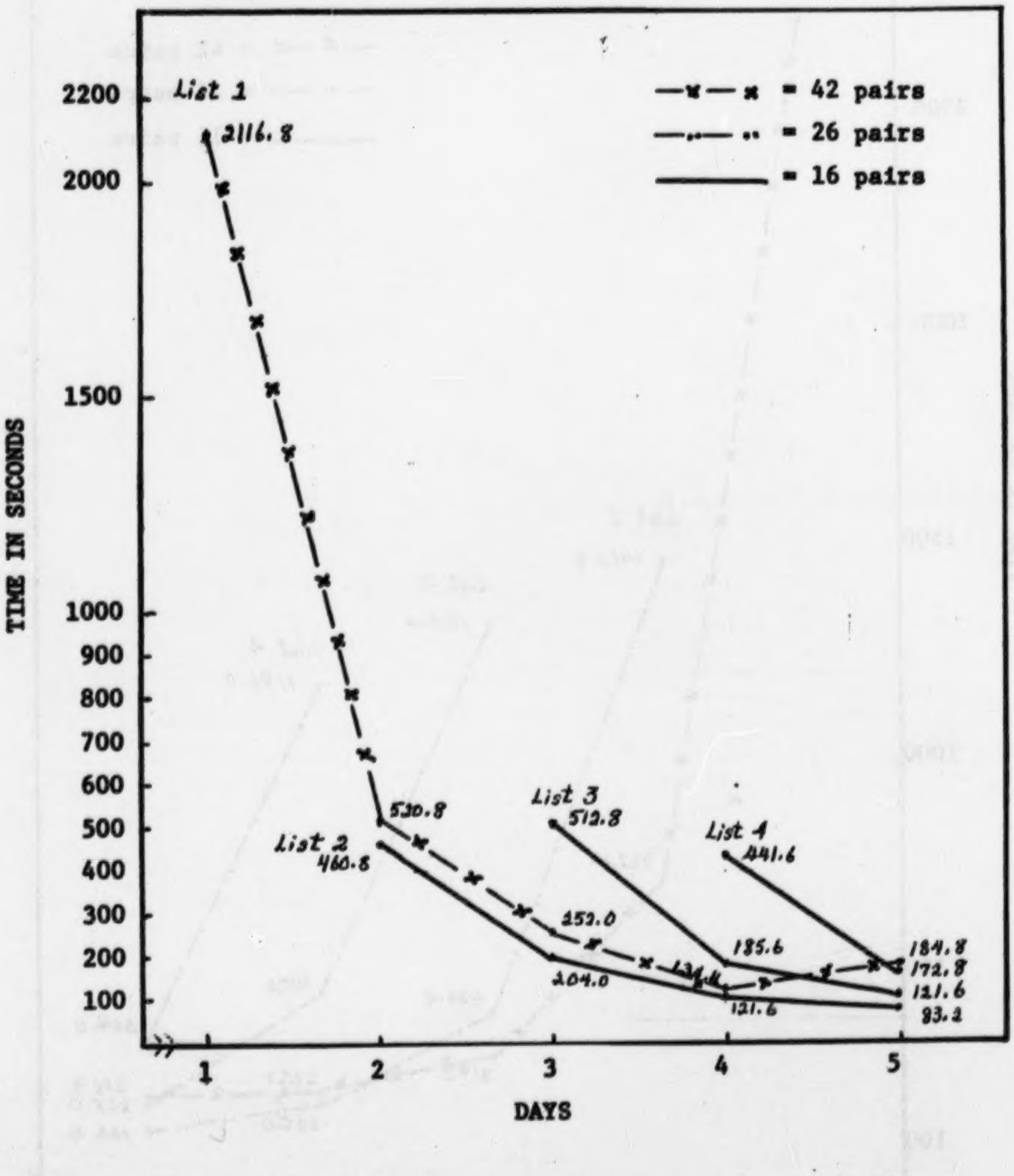
MEAN TIMES TO LEARN AND RELEARN PAIRS ON DAYS 1 TO 5 : GROUP 26 - 16

Time = (Trials - 1) x List Length x 4 seconds



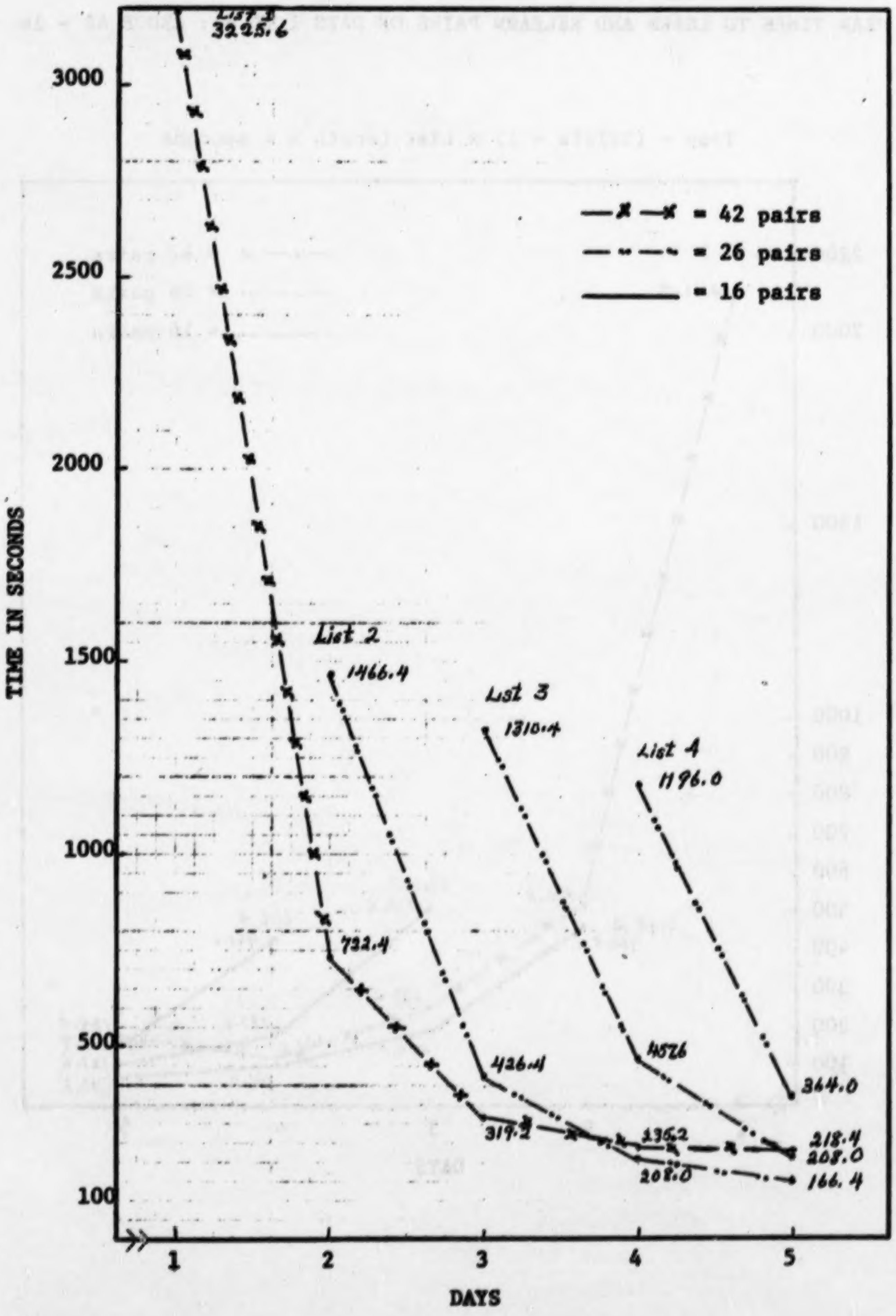
MEAN TIMES TO LEARN AND RELEARN PAIRS ON DAYS 1 TO 5 : GROUP 42 - 16

Time - (Trials - 1) x List Length x 4 seconds



MEAN TIMES TO LEARN AND RELEARN PAIRS ON DAYS 1 TO 5 : GROUP 42 - 26

Time = (Trials - 1) x List Length x 4 seconds

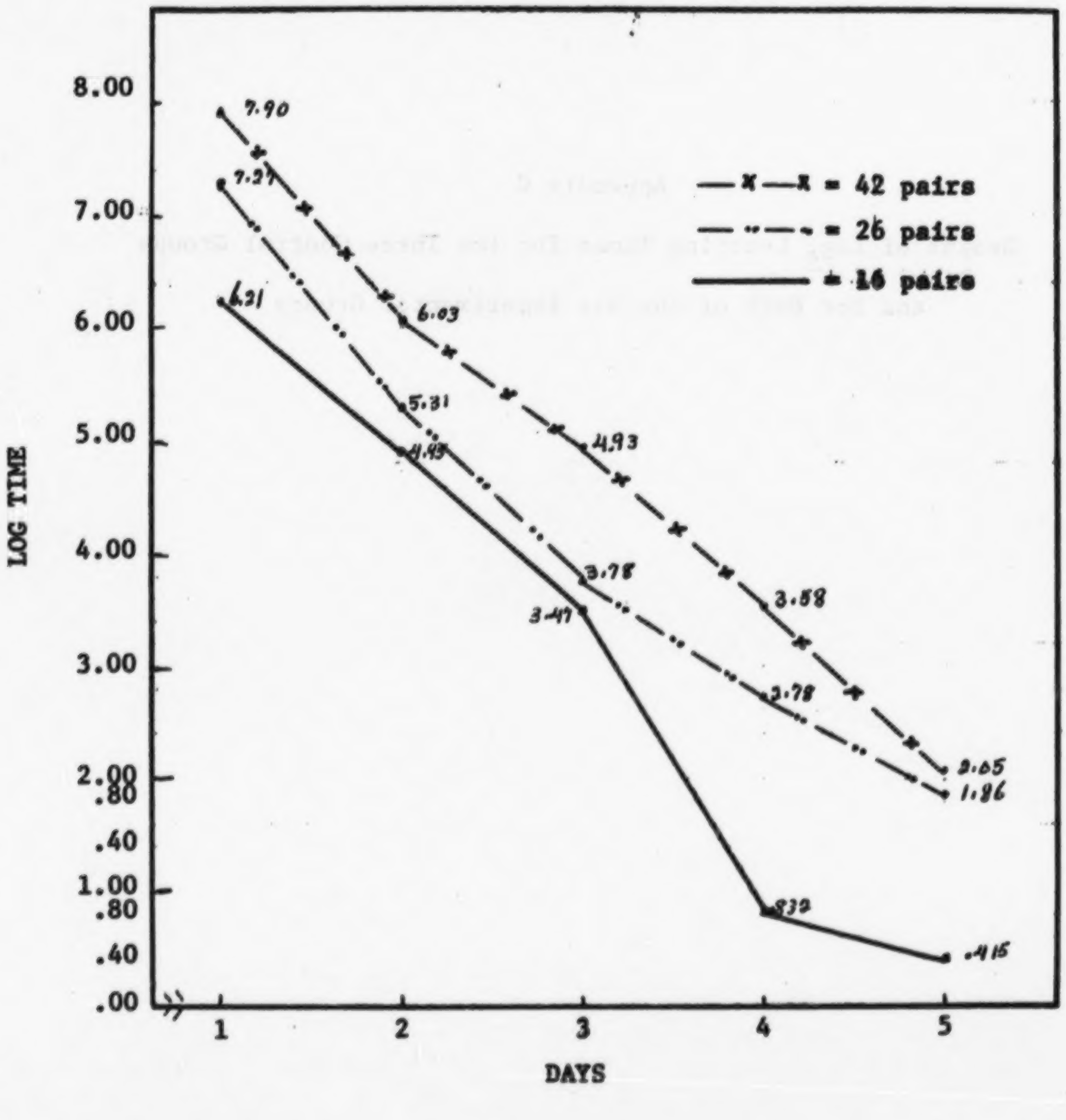


Appendix C

Graphs of Log$_e$ Learning Times for the Three Control Groups and for Each of the Six Experimental Groups

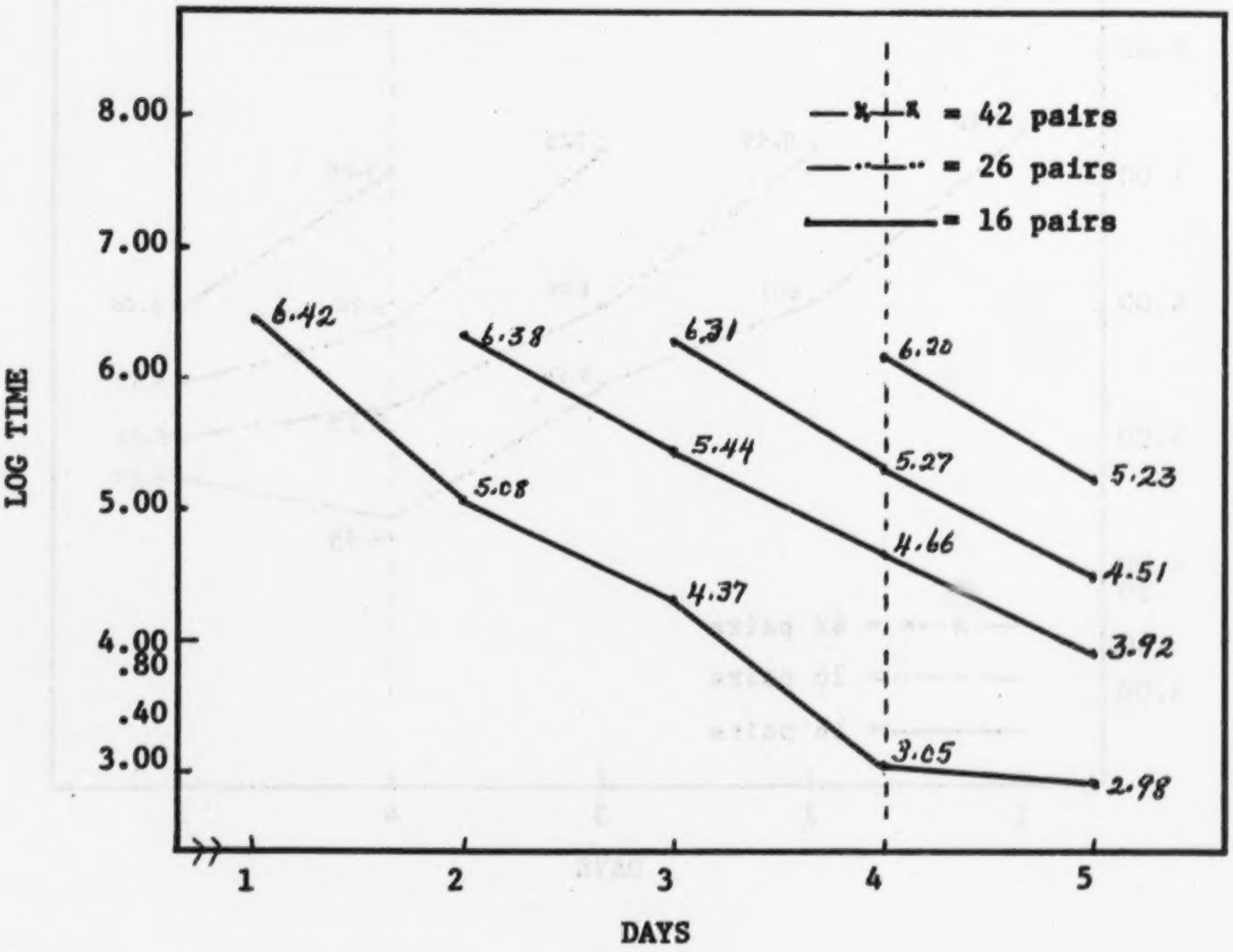
LOG$_e$ TOTAL TIME TO LEARN AND RELEARN OVER 5 DAYS : CONTROL GROUP

Time = (Trials - 1) x List Length x 4 seconds



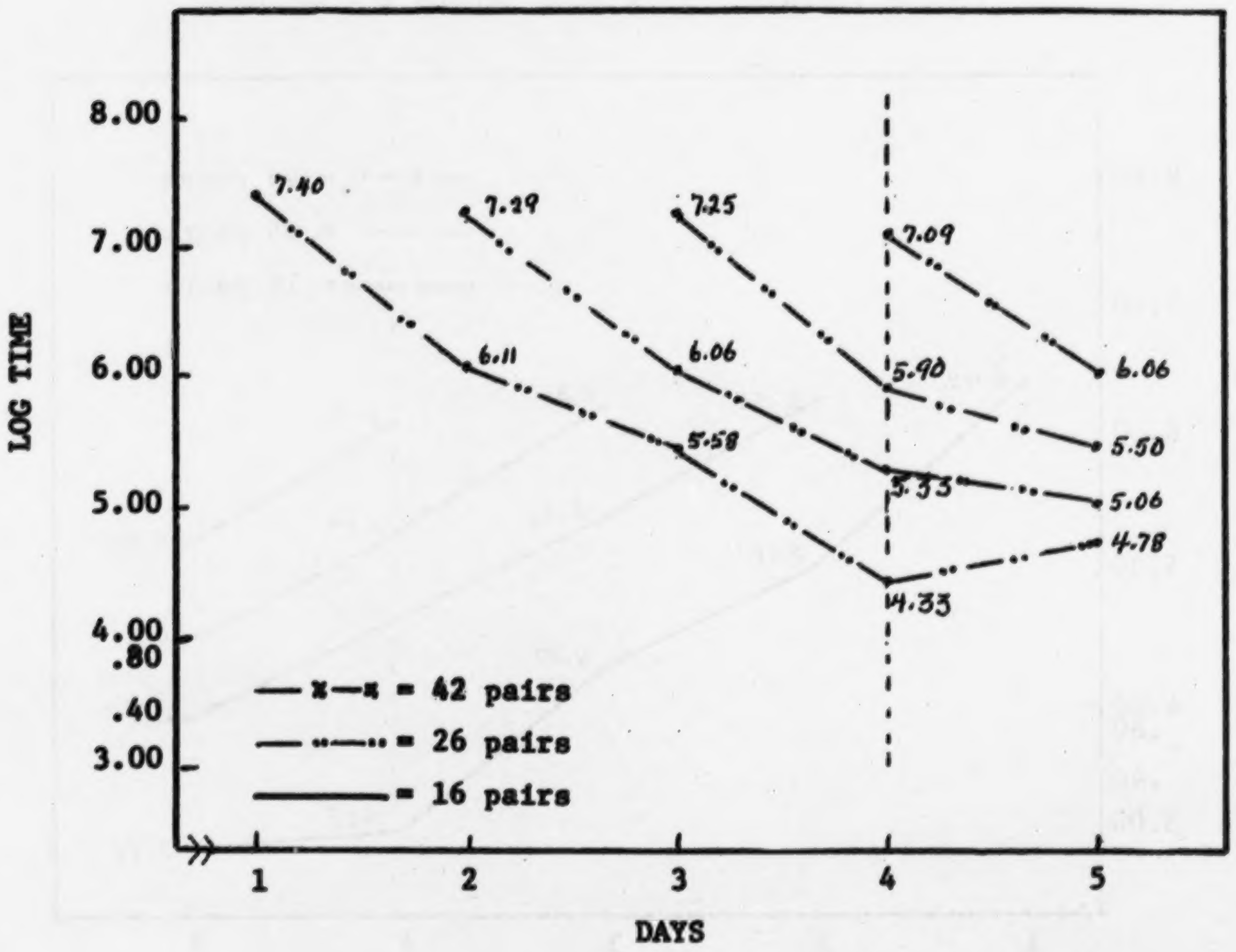
LOG$_e$ TOTAL TIME TO LEARN AND RELEARN OVER 5 DAYS : GROUP 16 - 16

Time = (Trials - 1) x List Length x 4 seconds



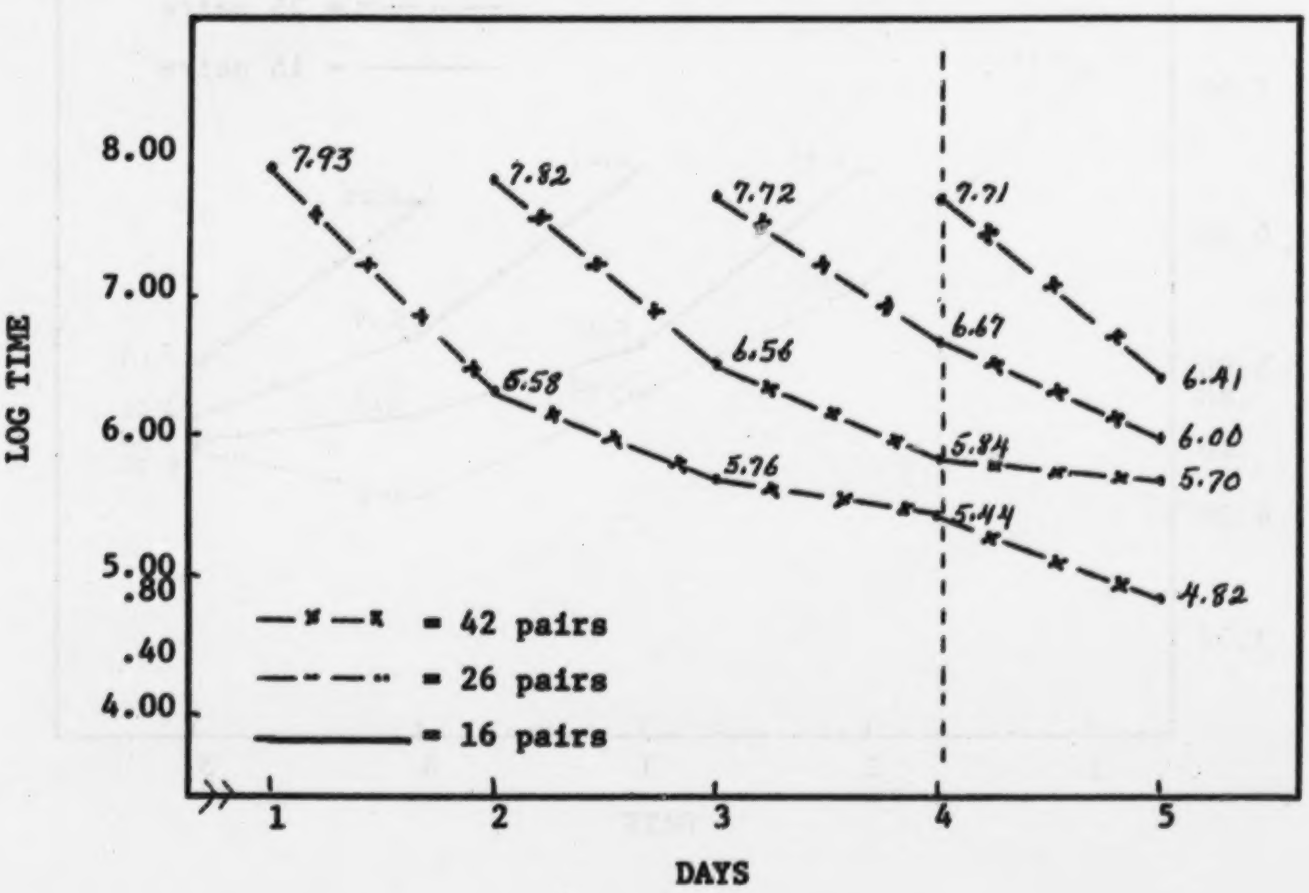
LOG$_e$ TOTAL TIME TO LEARN AND RELEARN OVER 5 DAYS : GROUP 26 - 26

Time = (Trials - 1) x List Length x 4 seconds



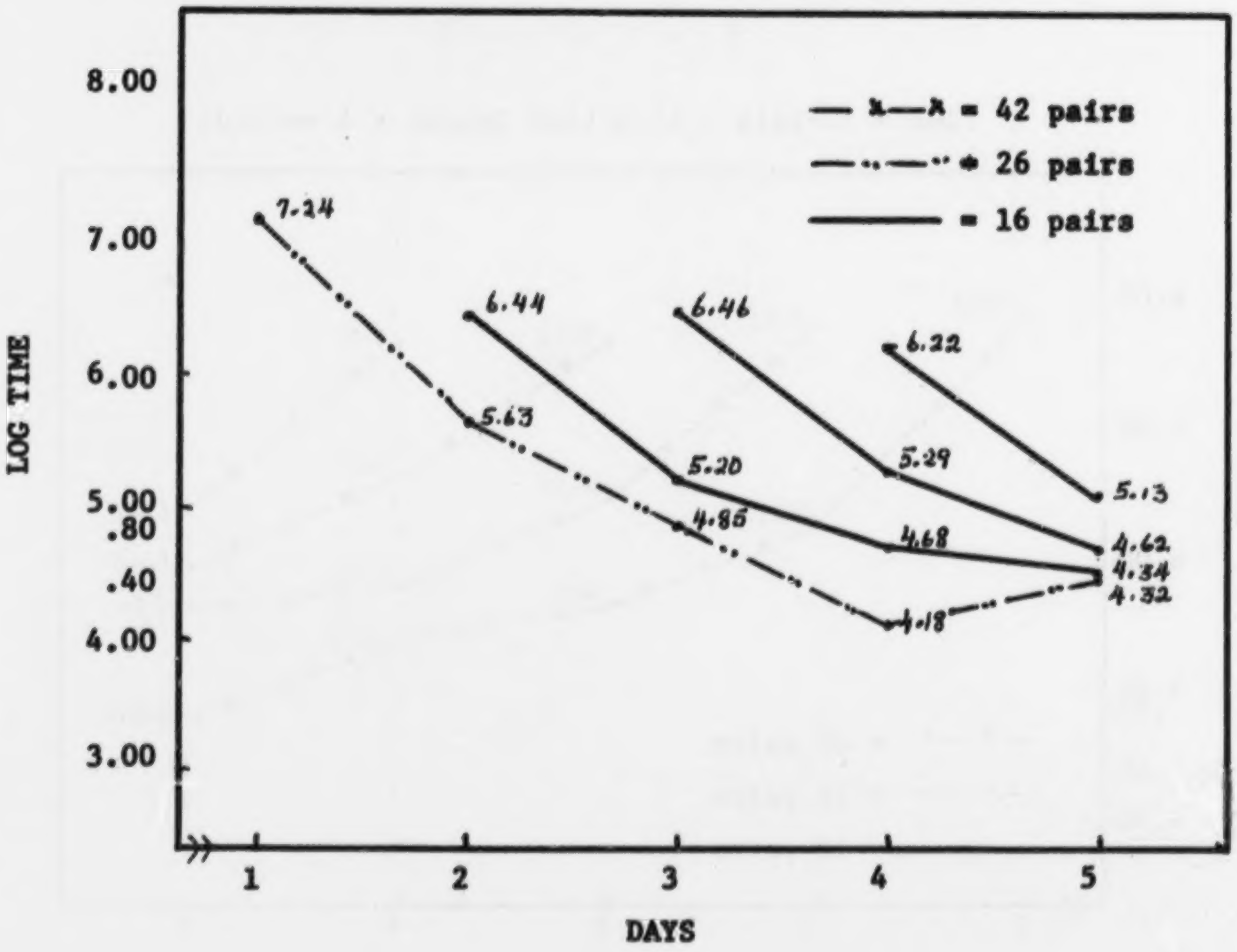
LOG$_e$ TOTAL TIME TO LEARN AND RELEARN OVER 5 DAYS : GROUP - 42 - 42

Time = (Trials - 1) x List Length x 4 seconds



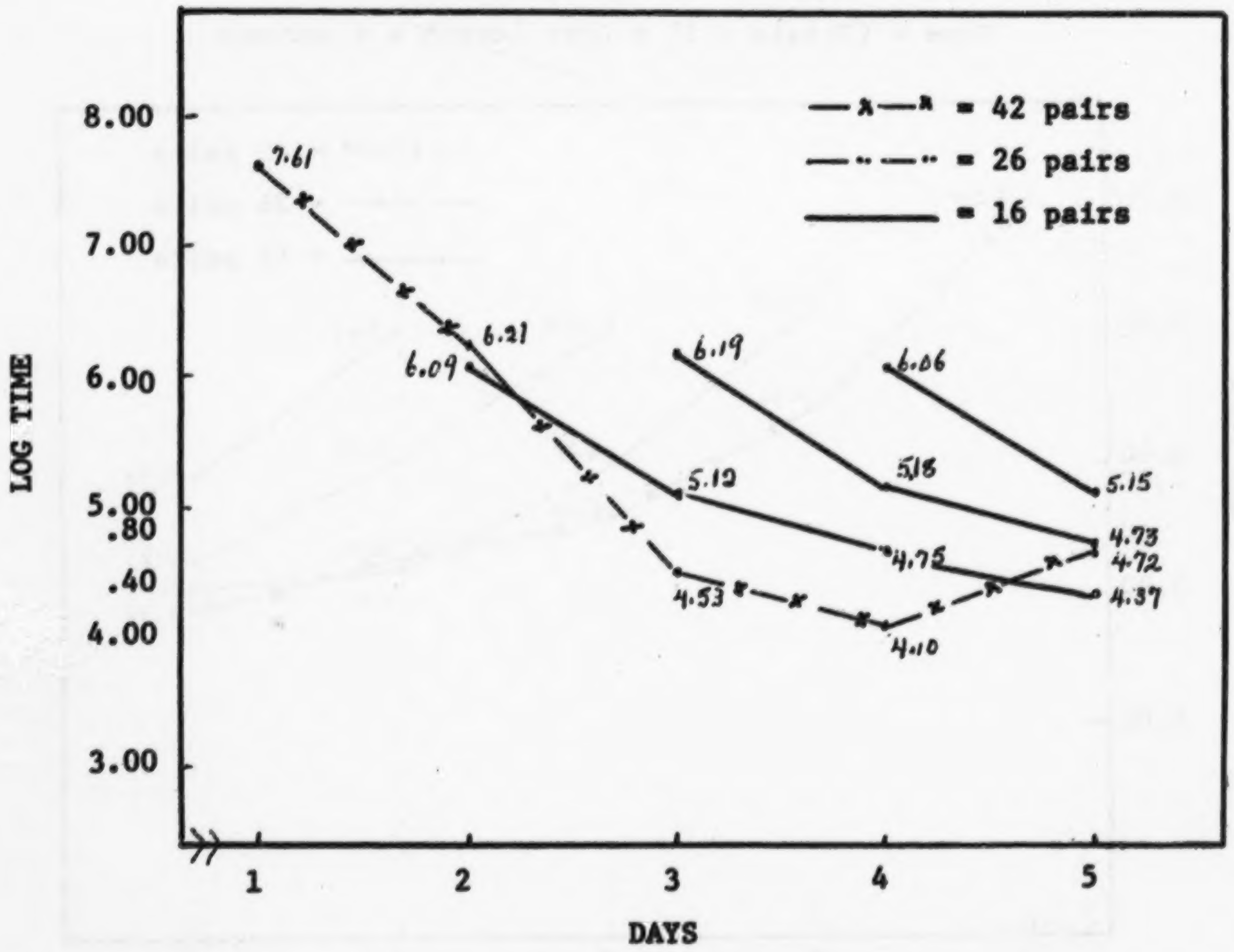
LOG$_e$ TOTAL TIME TO LEARN AND RELEARN OVER 5 DAYS : GROUP 26 - 16

Time = (Trials - 1) x List Length x 4 seconds



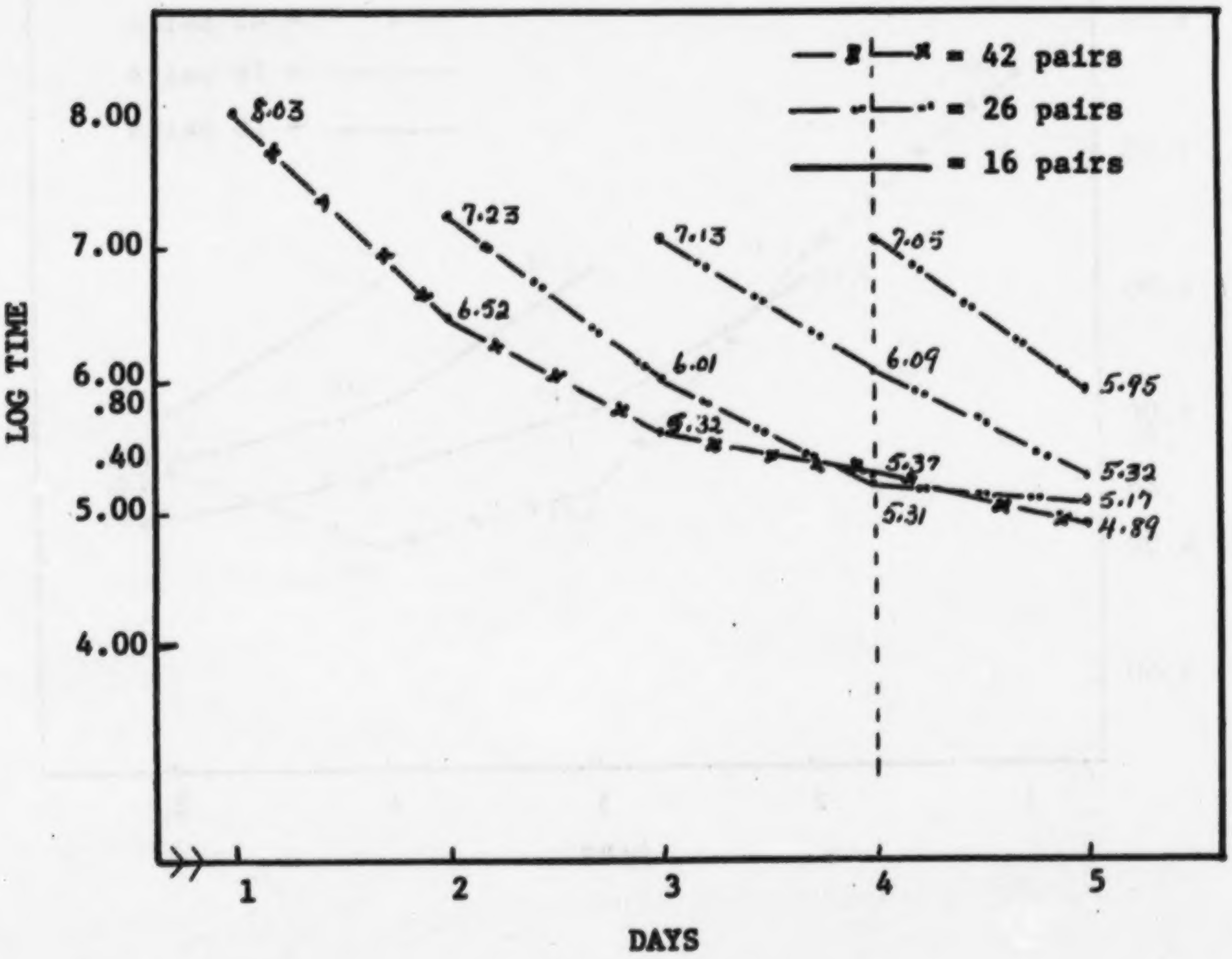
LOG$_e$ TOTAL TIME TO LEARN AND RELEARN OVER 5 DAYS : GROUP 42 - 16

Time = (Trials - 1) x List Length x 4 seconds



LOG$_e$ TOTAL TIME TO LEARN AND RELEARN OVER 5 DAYS : GROUP 42 - 26

Time = (Trials - 1) x List Length x 4 seconds

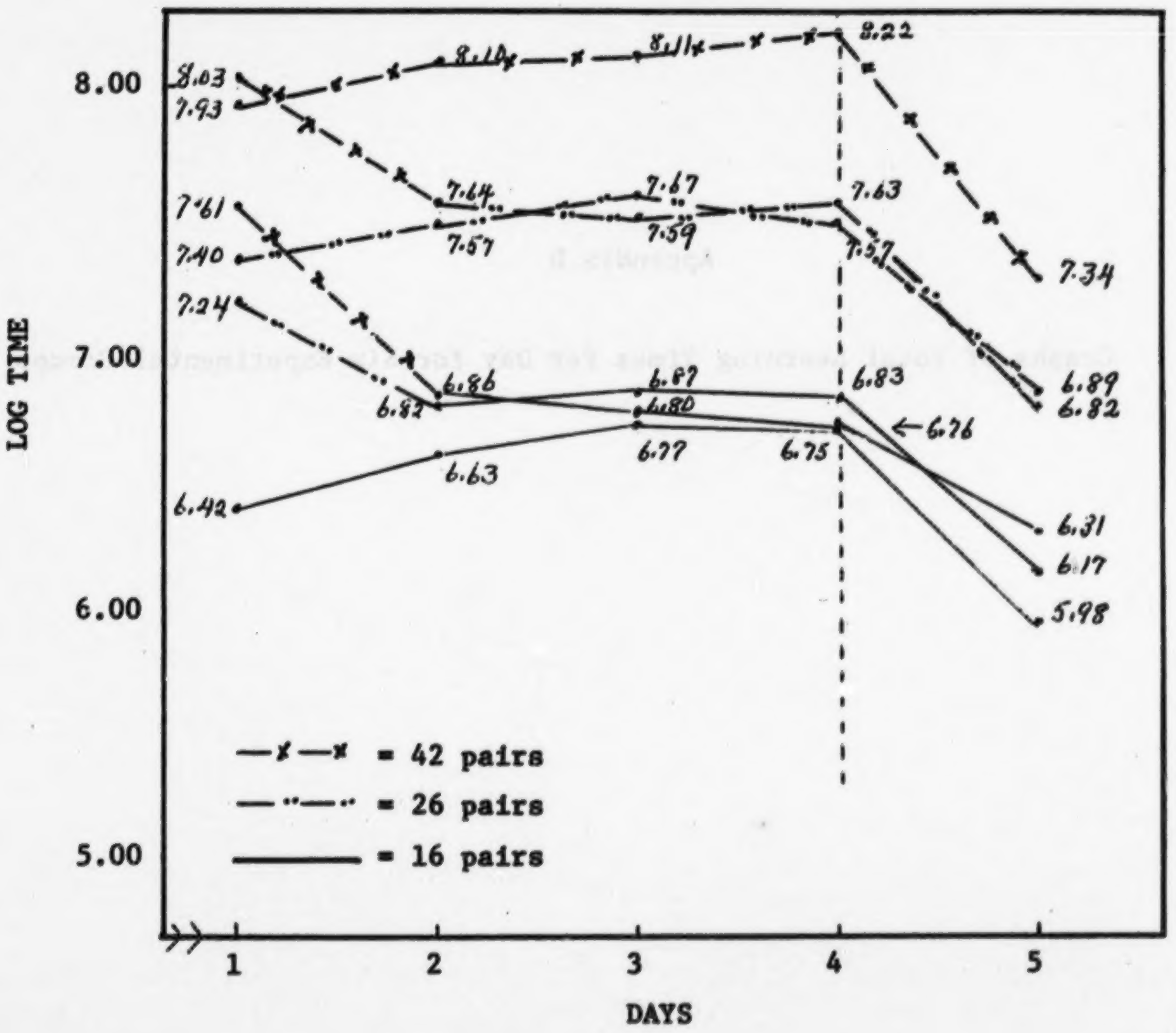


Appendix D

Graphs of Total Learning Times Per Day for Six Experimental Groups

LOG$_e$ TOTAL TIME TO LEARN AND RELEARN OVER 5 DAYS: ALL EXPERIMENTAL GROUPS

Time = (trials - 1) x List Length x 4 seconds

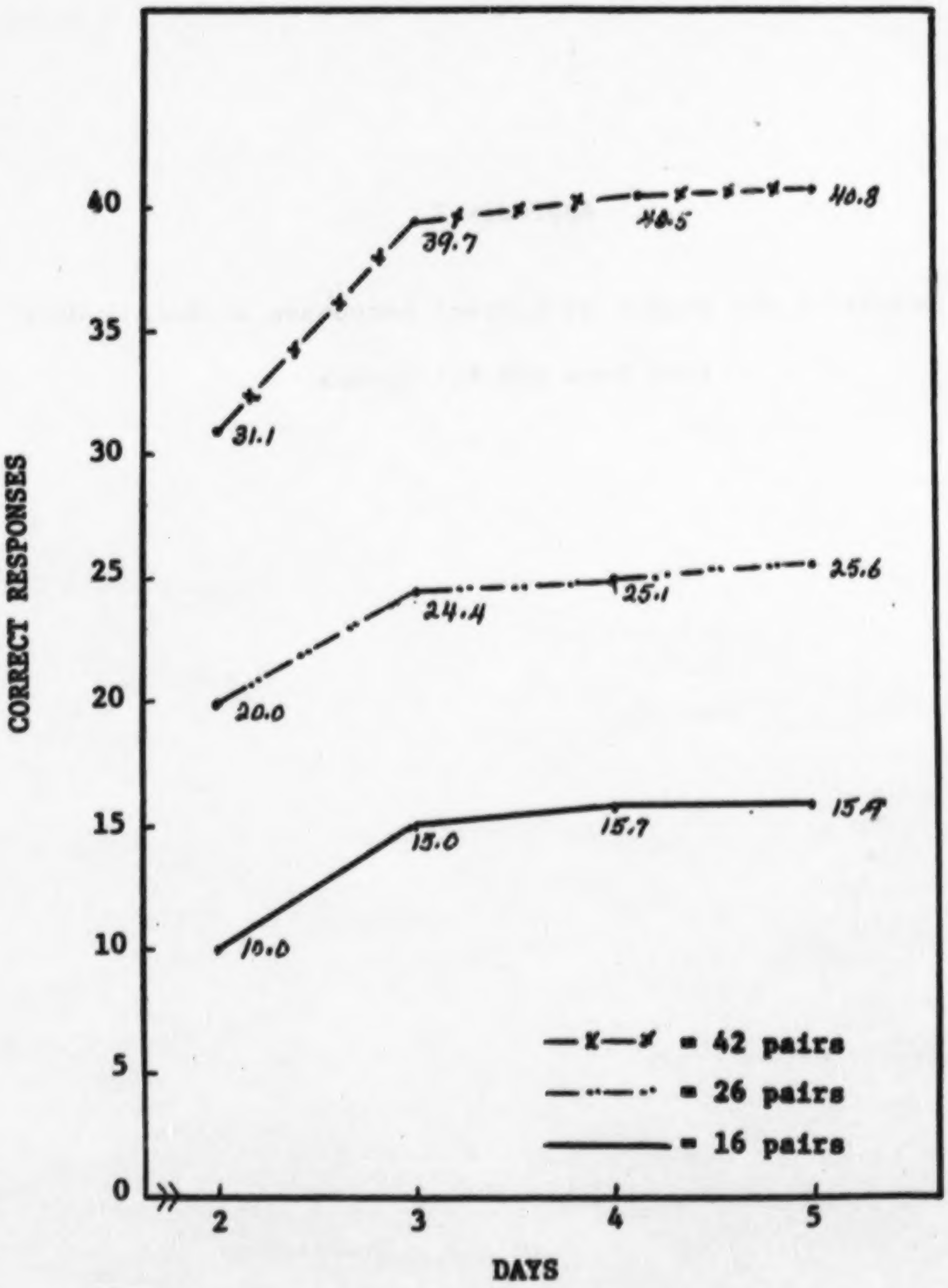


Appendix E

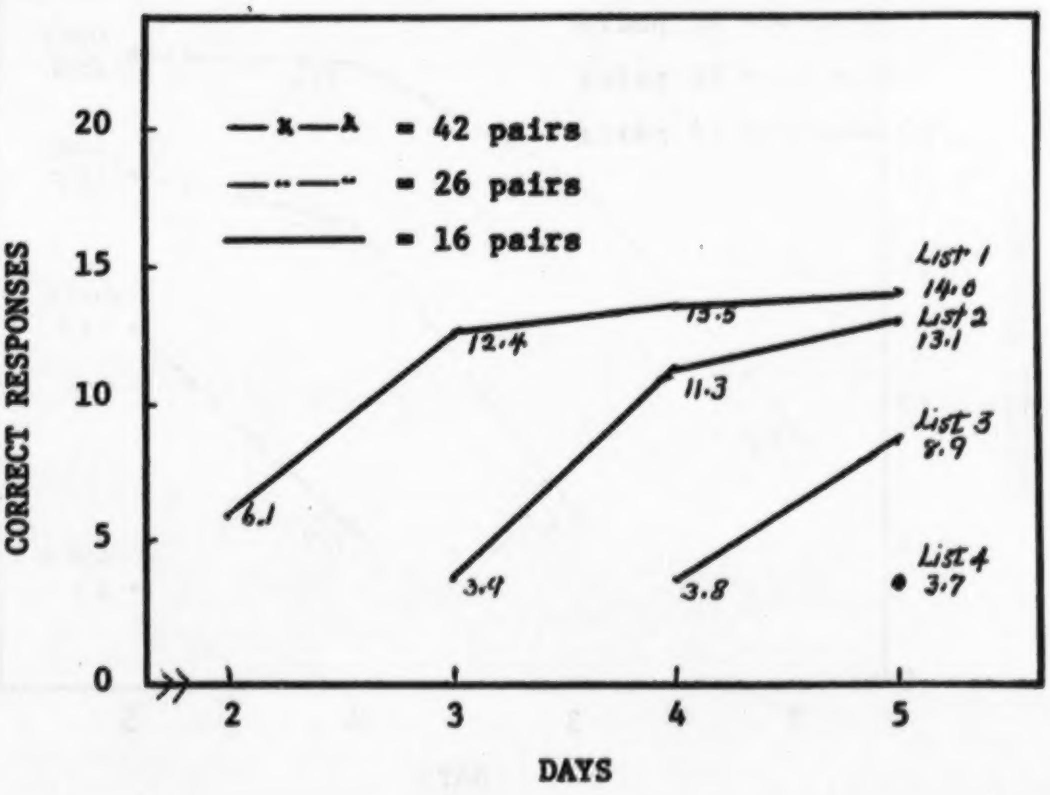
Graphs of the Number of Correct Responses at Recall Over Four Days for All Groups

NUMBER OF ITEMS CORRECTLY ANTICIPATED ON TRIAL ONE,

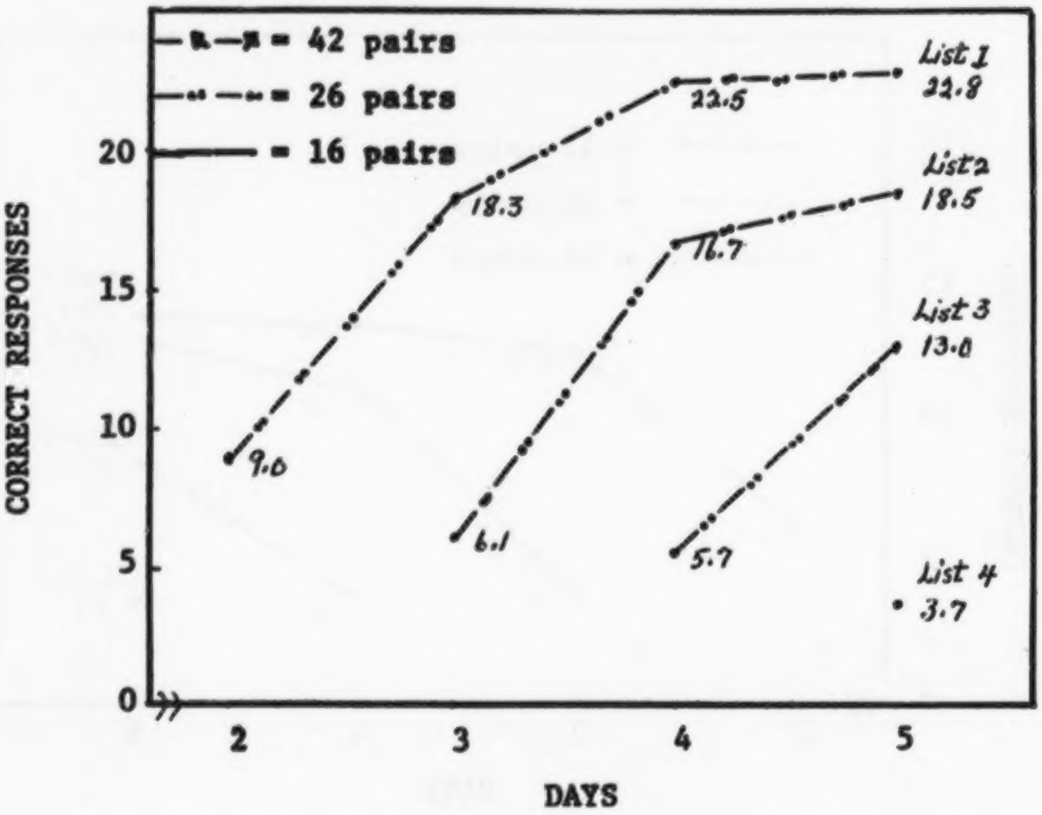
DAYS TWO THROUGH FIVE, ALL LISTS: CONTROL GROUP



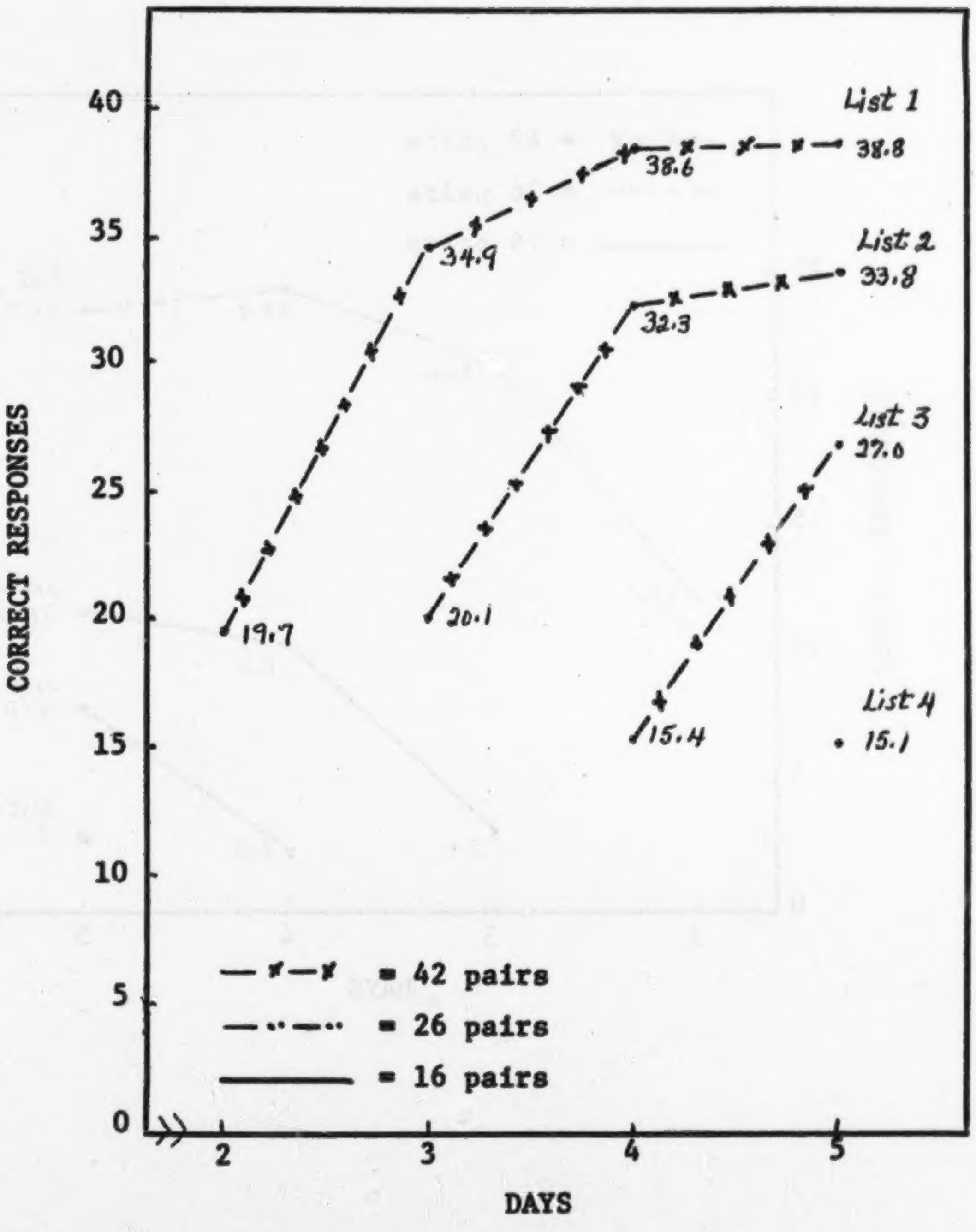
NUMBER OF ITEMS CORRECTLY ANTICIPATED ON TRIAL ONE, DAYS ONE THROUGH FIVE, ALL LISTS: GROUP 16 - 16



NUMBER OF ITEMS CORRECTLY ANTICIPATED ON TRIAL ONE,
DAYS TWO THROUGH FIVE, ALL LISTS: GROUP 26 - 26

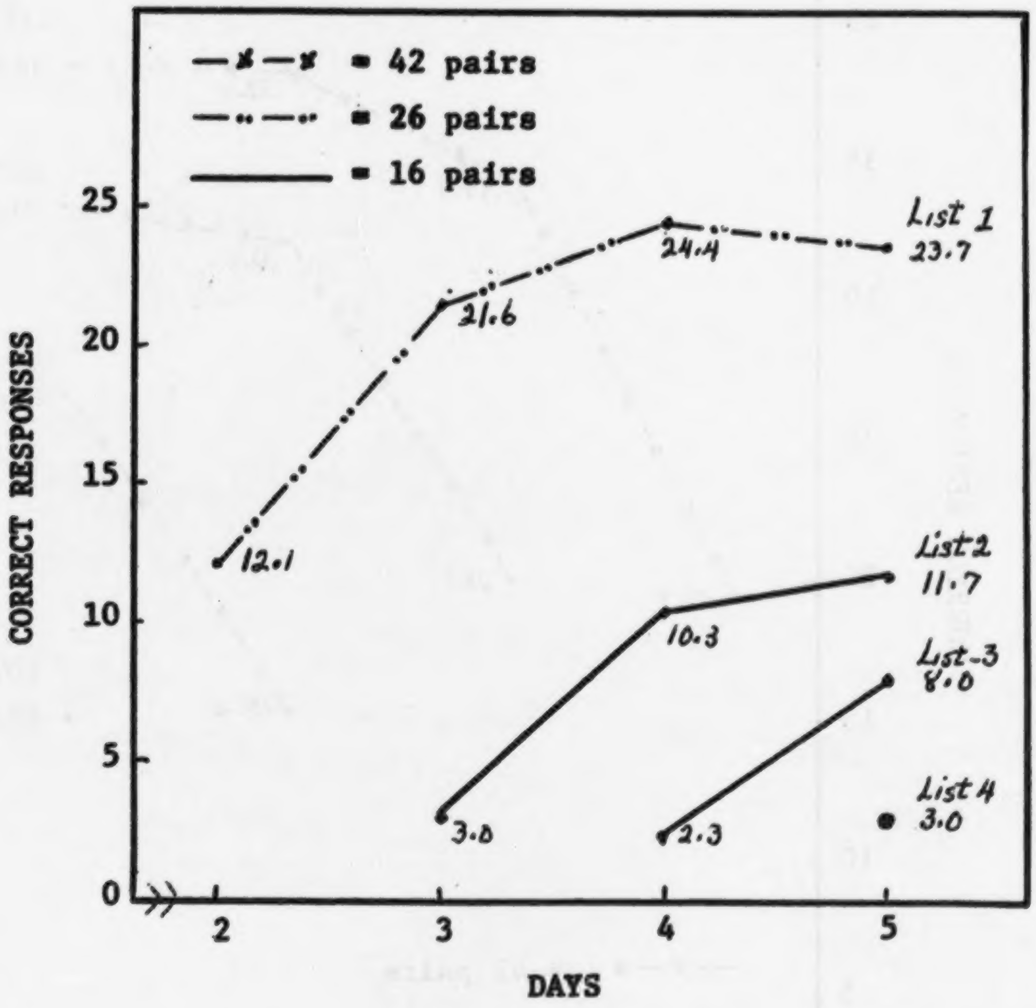


NUMBER OF ITEMS CORRECTLY ANTICIPATED ON TRIAL ONE,
DAYS TWO THROUGH FIVE, ALL LISTS: GROUP 42 - 42

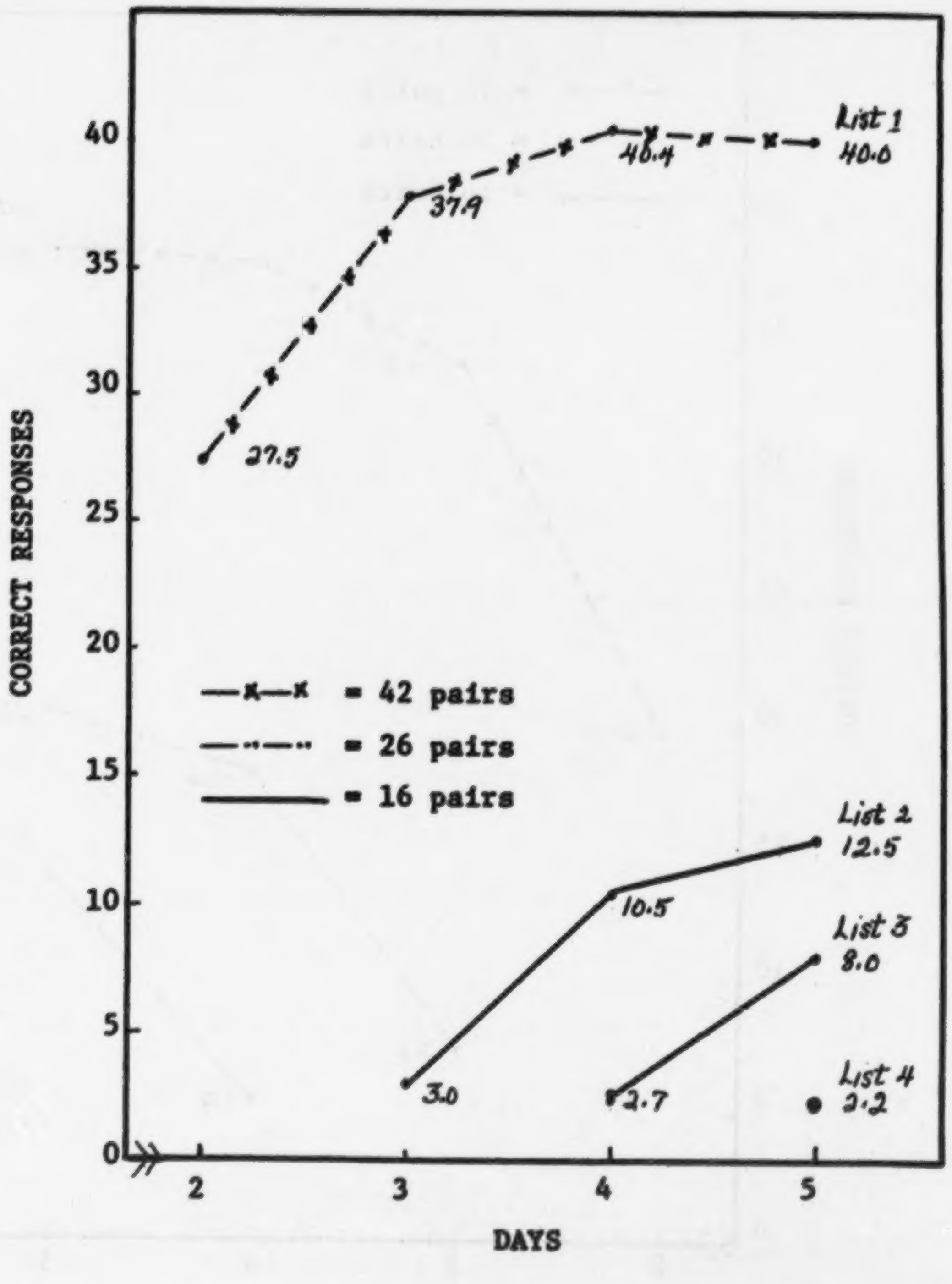


NUMBER OF ITEMS CORRECTLY ANTICIPATED ON TRIAL ONE,

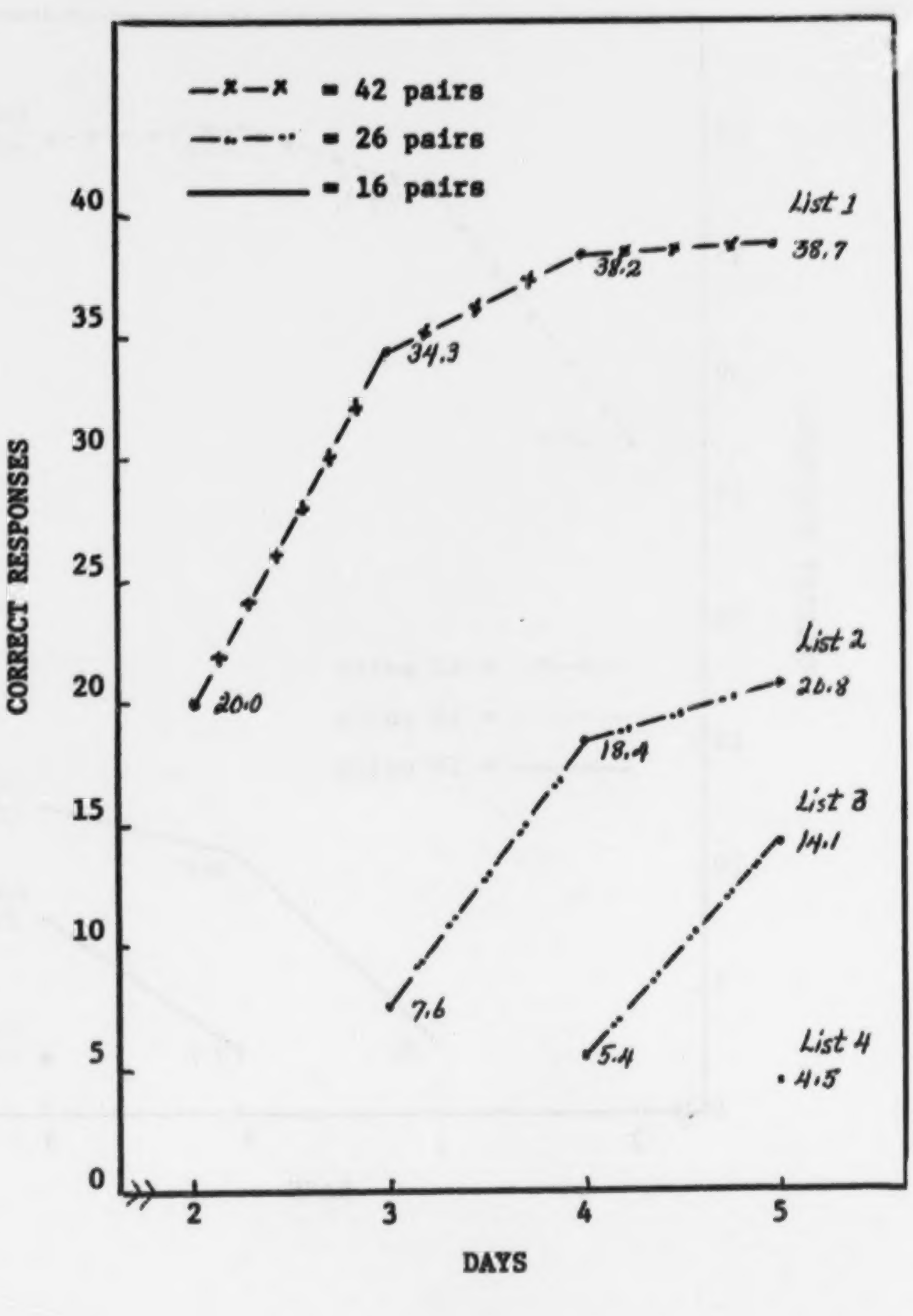
DAYS TWO THROUGH FIVE, ALL LISTS: GROUP 26 - 16



NUMBER OF ITEMS CORRECTLY ANTICIPATED ON TRIAL ONE, DAYS TWO THROUGH FIVE, ALL LISTS: GROUP 42 - 16



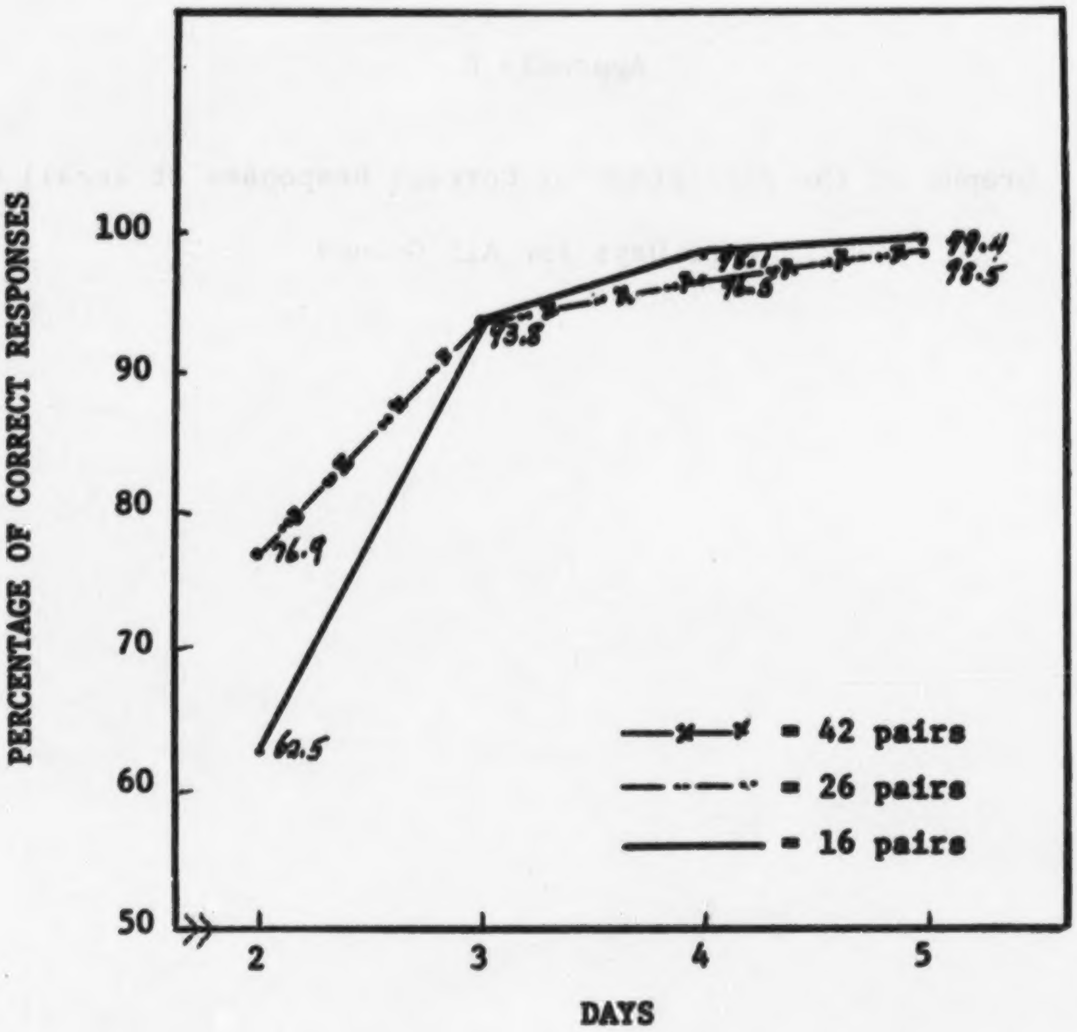
NUMBER OF ITEMS CORRECTLY ANTICIPATED ON TRIAL ONE,
DAYS TWO THROUGH FIVE, ALL LISTS: GROUP 42 - 26



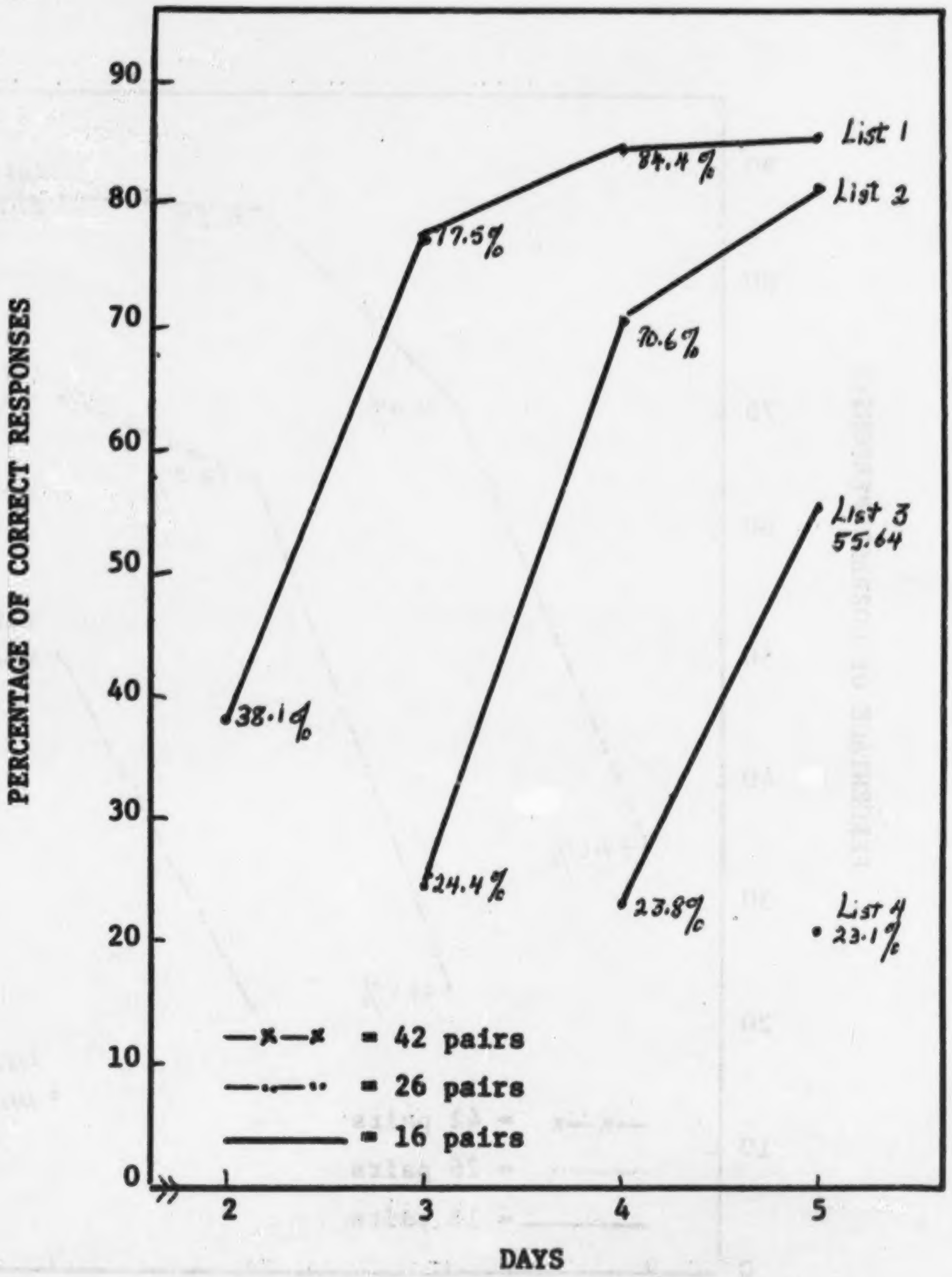
Appendix F

Graphs of the Percentage of Correct Responses at Recall Over Four Days for All Groups

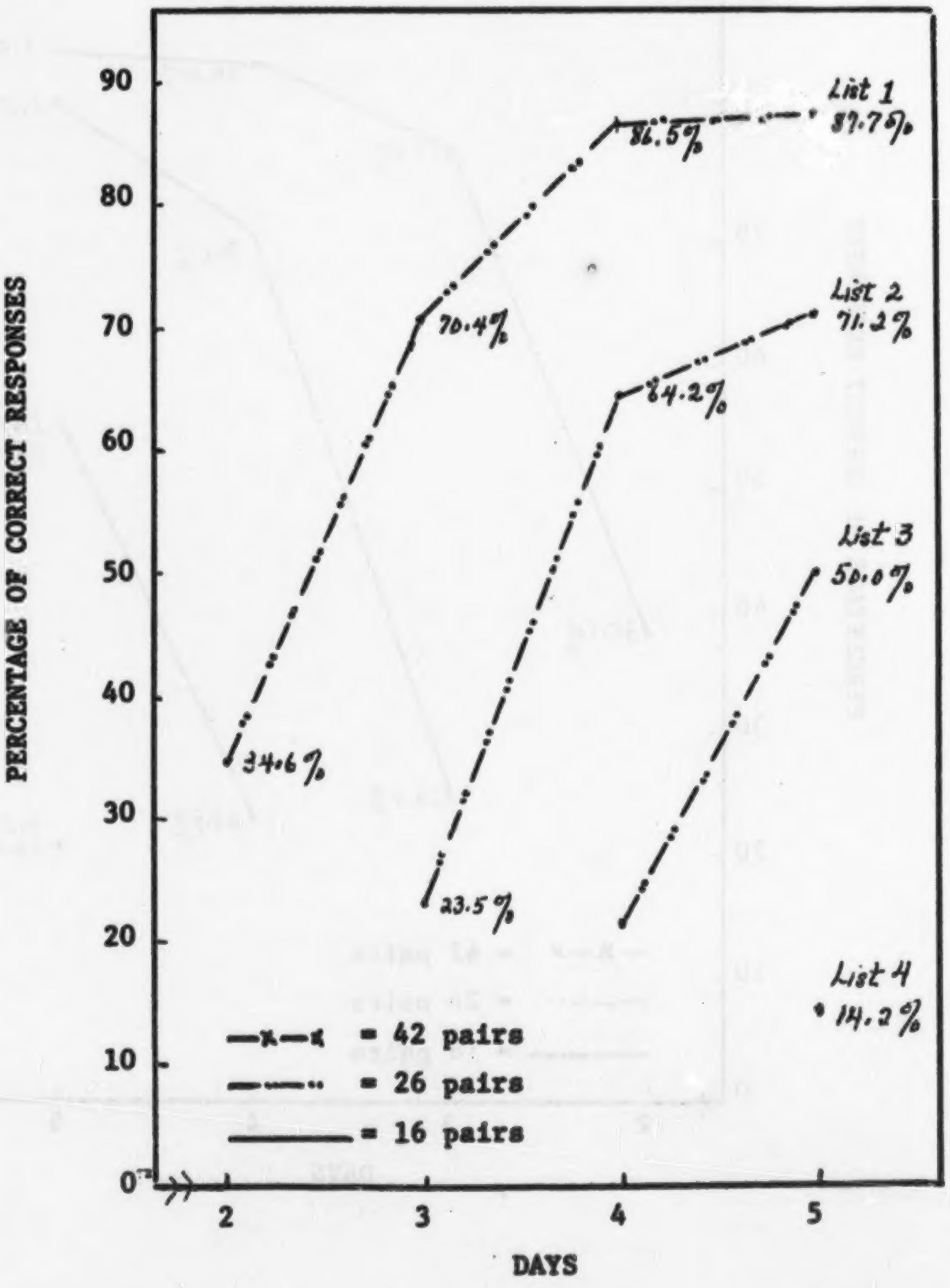
PERCENT OF ITEMS RECALLED ON TRIAL ONE, DAYS TWO THROUGH FIVE, ALL LISTS: CONTROL GROUP



PERCENT OF ITEMS RECALLED ON TRIAL ONE, DAYS TWO THROUGH FIVE, ALL LISTS:
GROUP 16 - 16

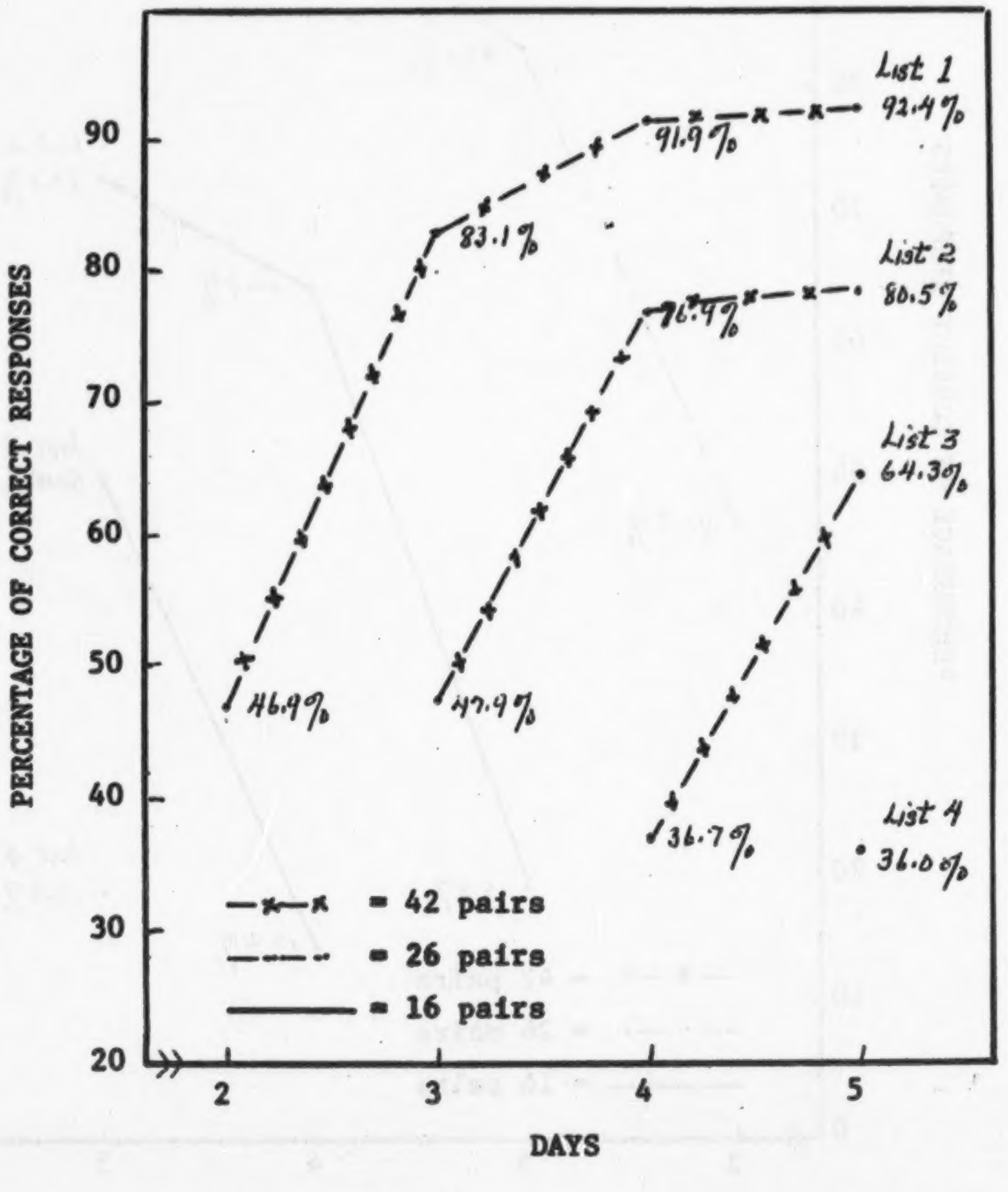


PERCENT OF ITEMS RECALLED ON TRIAL ONE, DAYS TWO THROUGH FIVE, ALL LISTS: GROUP 26 - 26



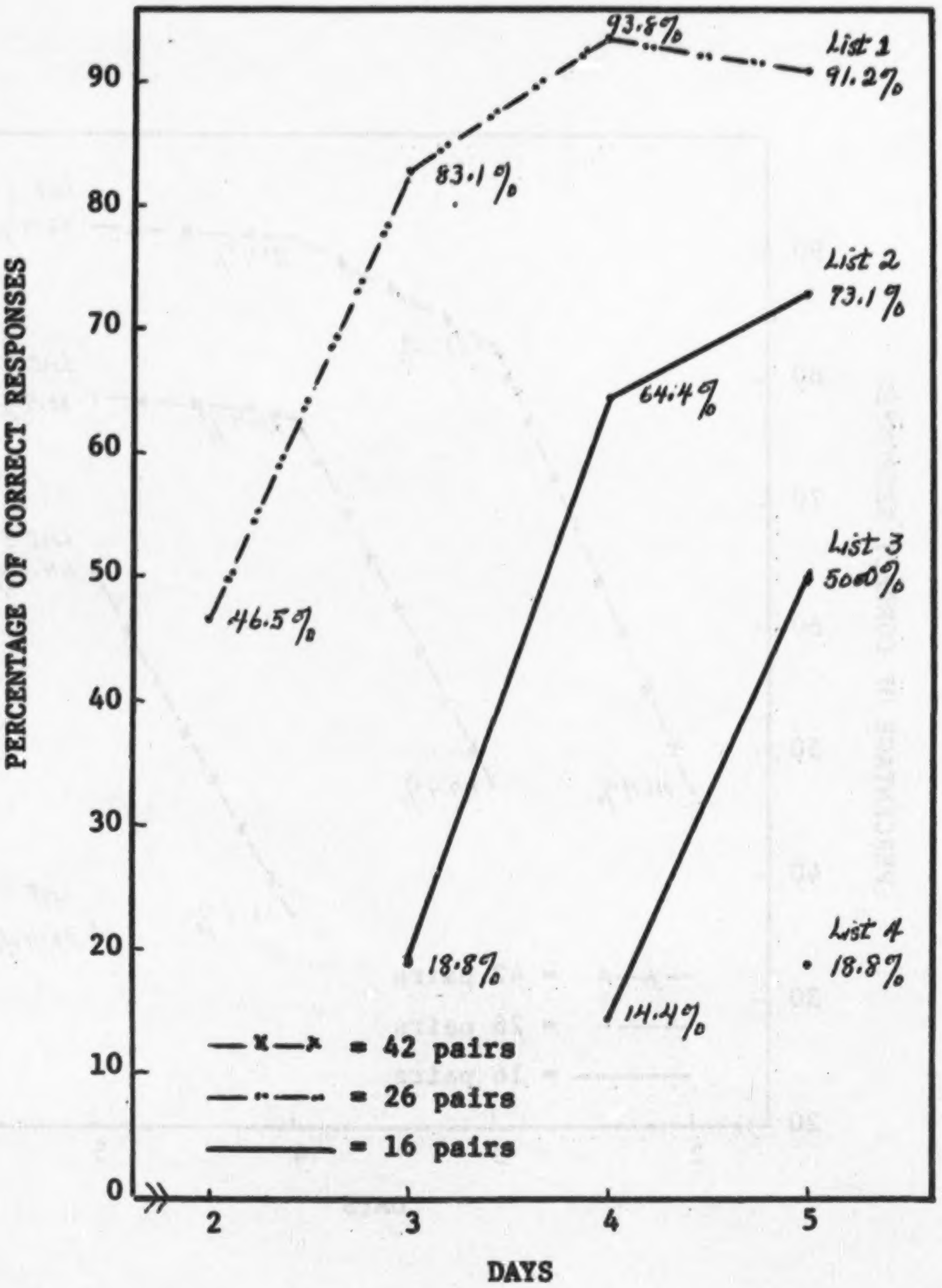
PERCENT OF ITEMS RECALLED ON TRIAL ONE, DAYS TWO THROUGH FIVE, ALL LISTS:

GROUP 42 - 42

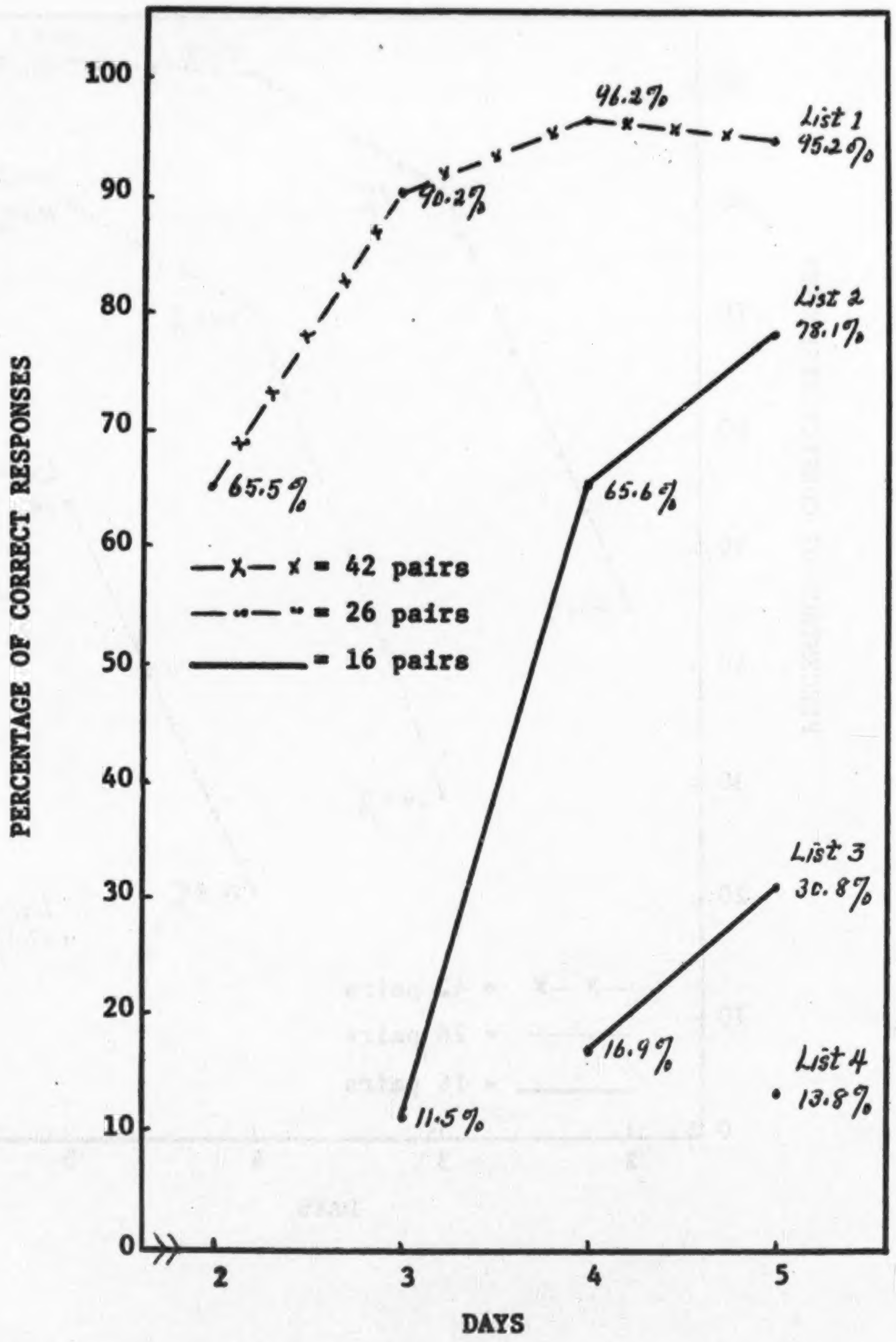


PERCENT OF ITEMS RECALLED ON TRIAL ONE, DAYS TWO THROUGH FIVE, ALL LISTS:

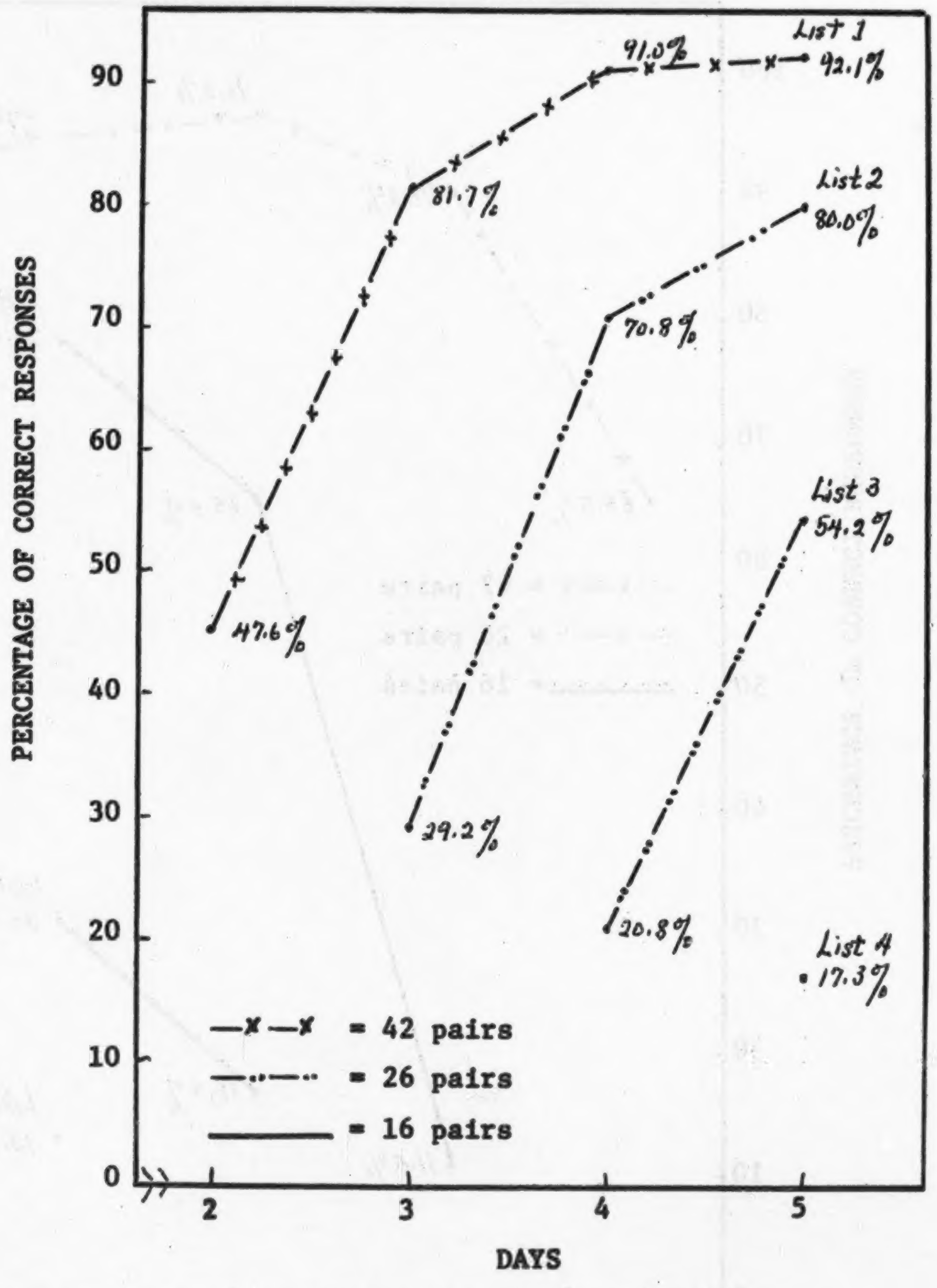
GROUP 26 - 16



PERCENT OF ITEMS RECALLED ON TRIAL ONE, DAYS TWO THROUGH FIVE, ALL LISTS: GROUP 42 - 16



PERCENT OF ITEMS RECALLED ON TRIAL ONE, DAYS TWO THROUGH FIVE, ALL LISTS: GROUP 42 - 26

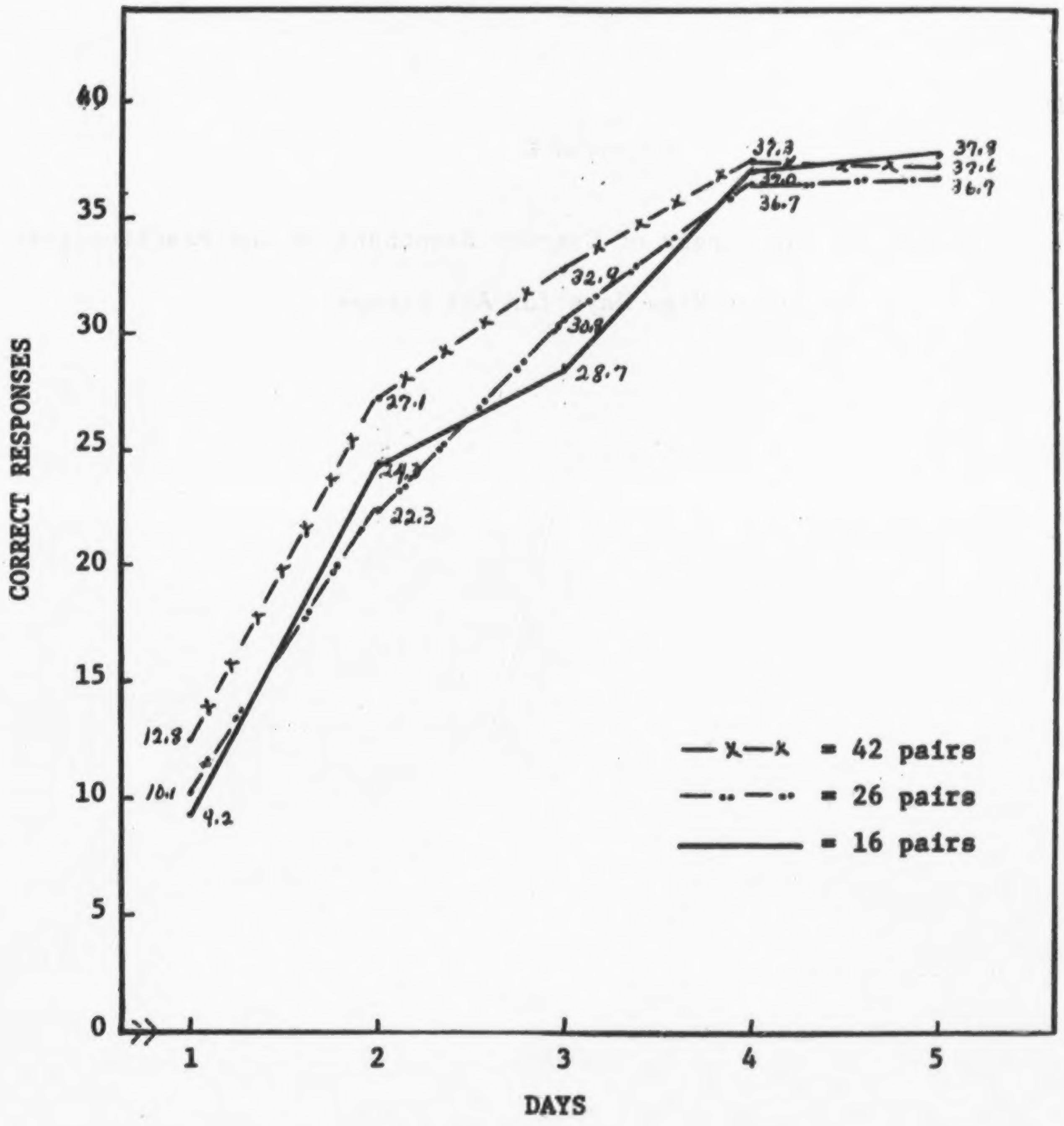


Appendix G

Graphs of the Number of Correct Responses to the Practice List Over Five Days for All Groups

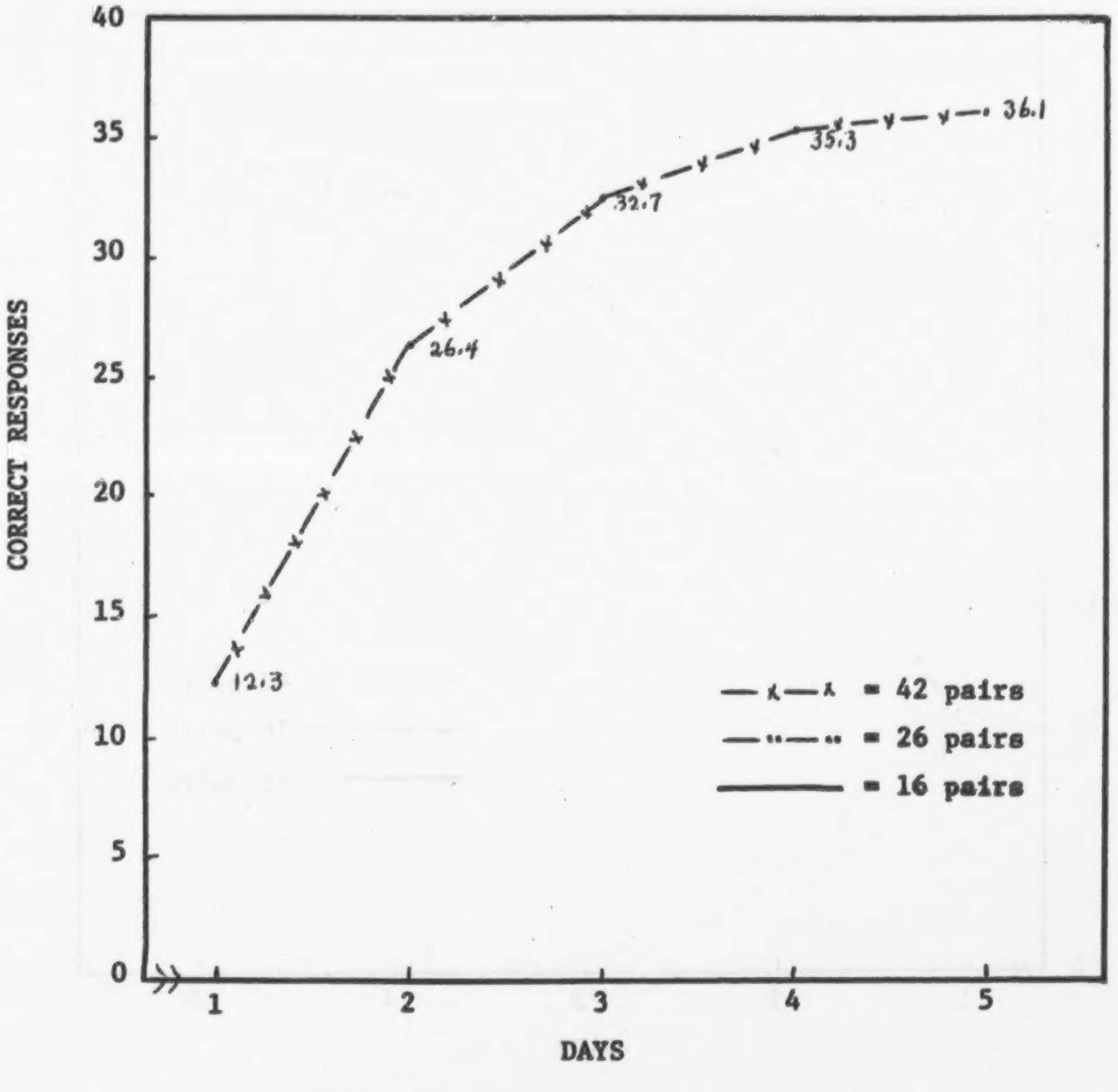
NUMBER OF CORRECT RESPONSES OVER 5 TRIALS ON PRACTICE LIST,

DAYS ONE THROUGH FIVE : CONTROL GROUP



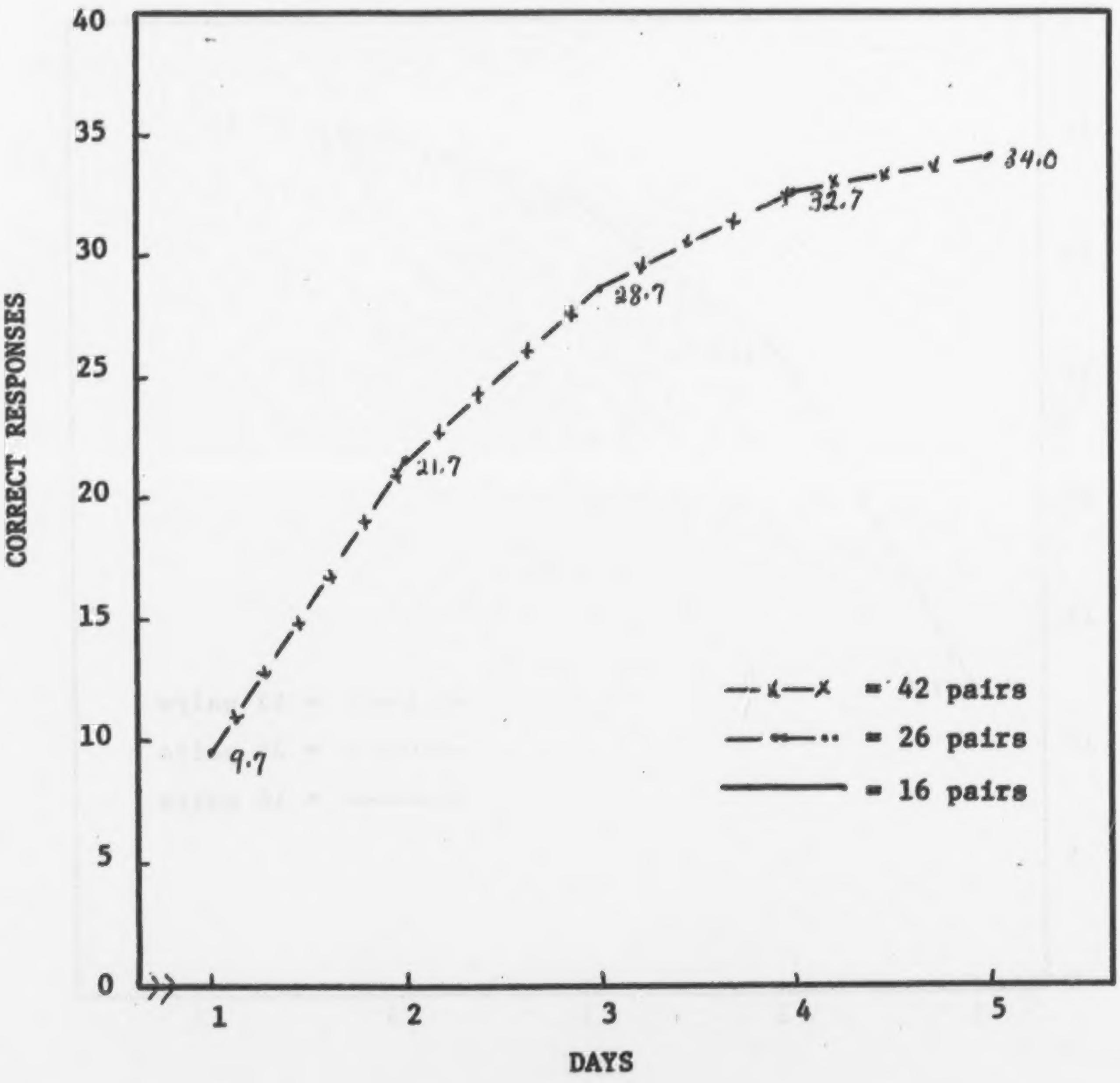
NUMBER OF CORRECT RESPONSES OVER 5 TRIALS ON PRACTICE LIST,

DAYS ONE THROUGH FIVE : GROUP 16 - 16

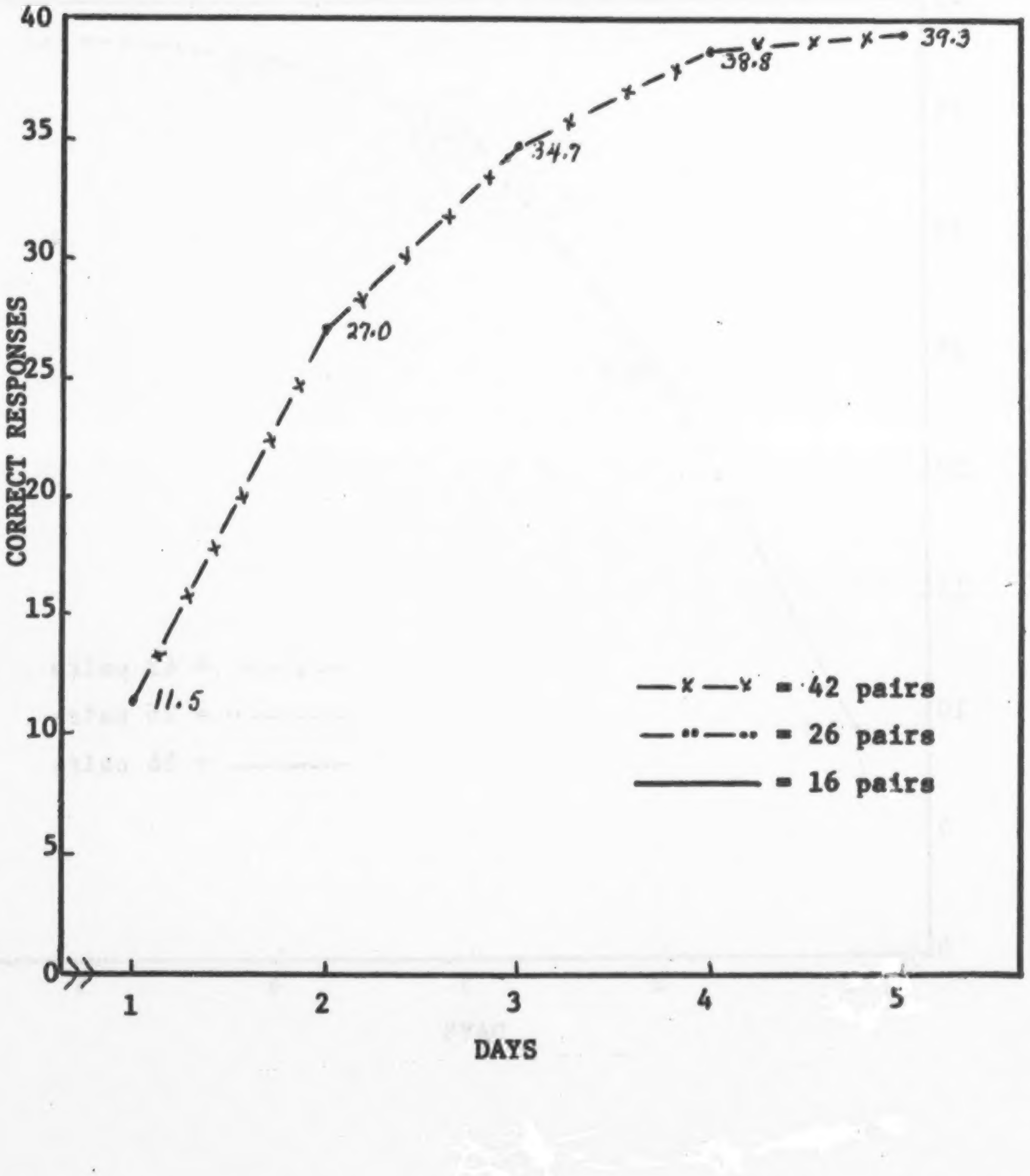


NUMBER OF CORRECT RESPONSES OVER 5 TRIALS ON PRACTICE LIST,

DAYS ONE THROUGH FIVE : GROUP 26 - 26

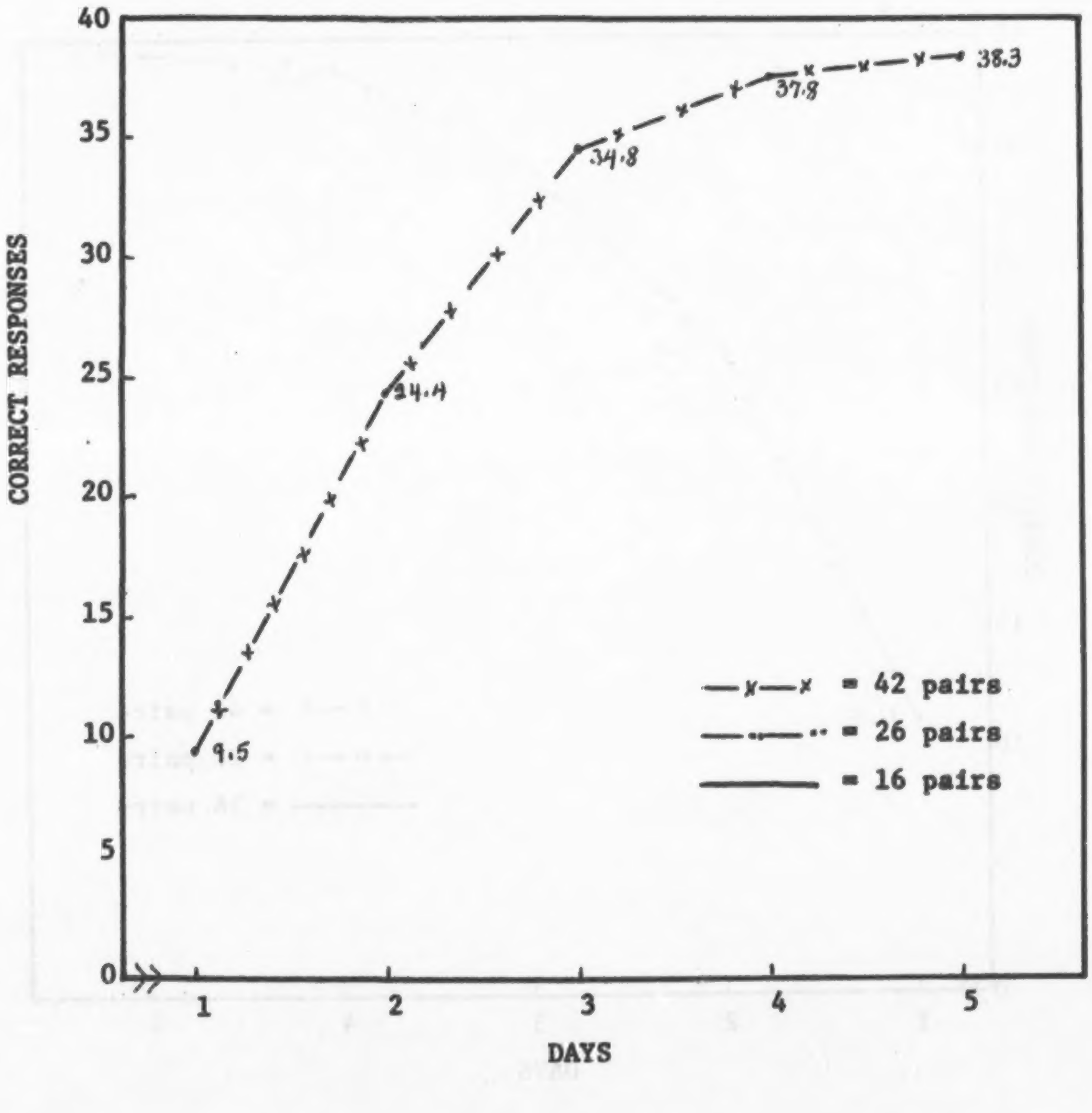


NUMBER OF CORRECT RESPONSES OVER 5 TRIALS ON PRACTICE LIST,
DAYS ONE THROUGH FIVE : GROUP 42 - 42



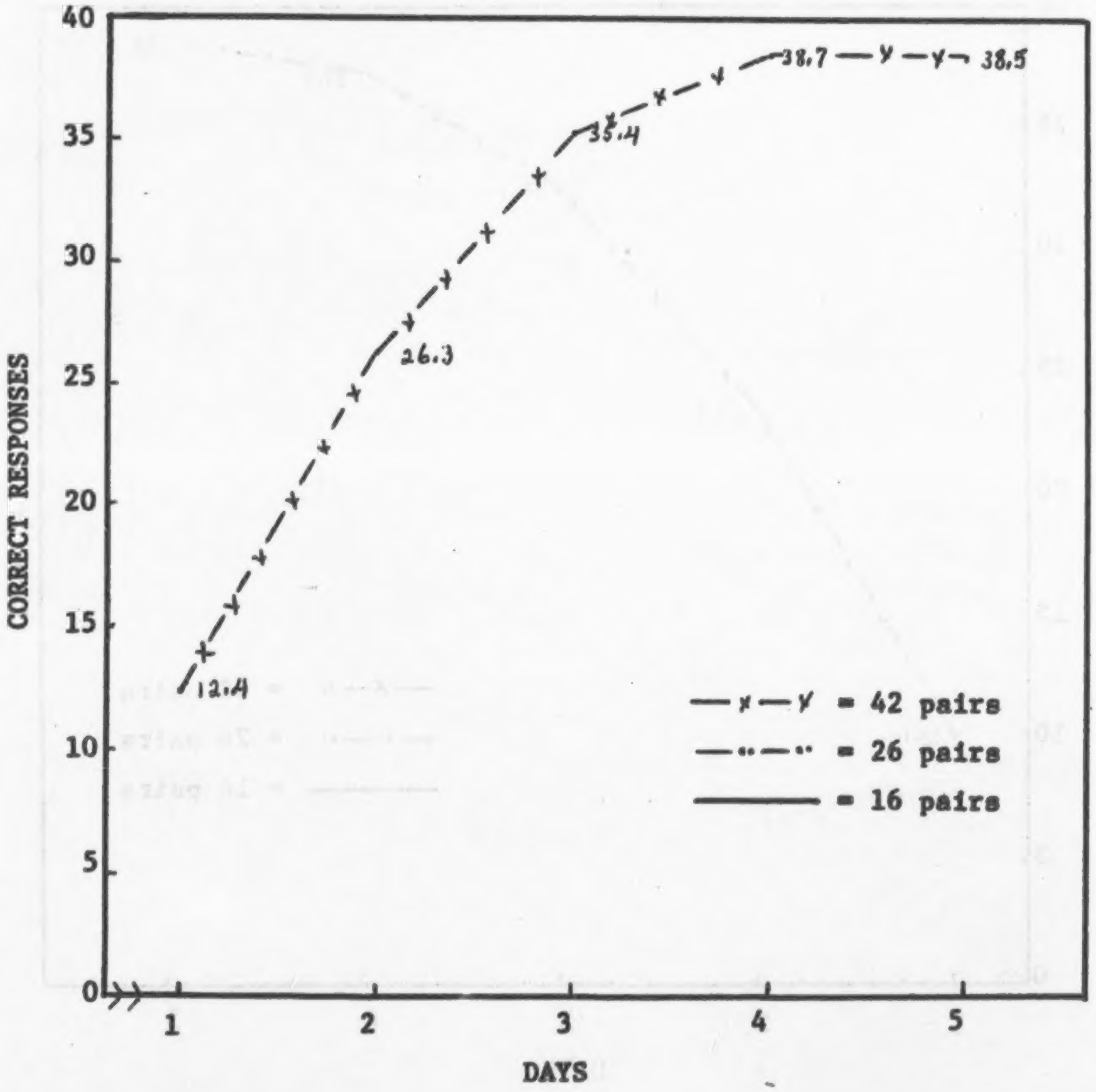
NUMBER OF CORRECT RESPONSES OVER 5 TRIALS ON PRACTICE LIST,

DAYS ONE THROUGH FIVE : GROUP 26 - 16



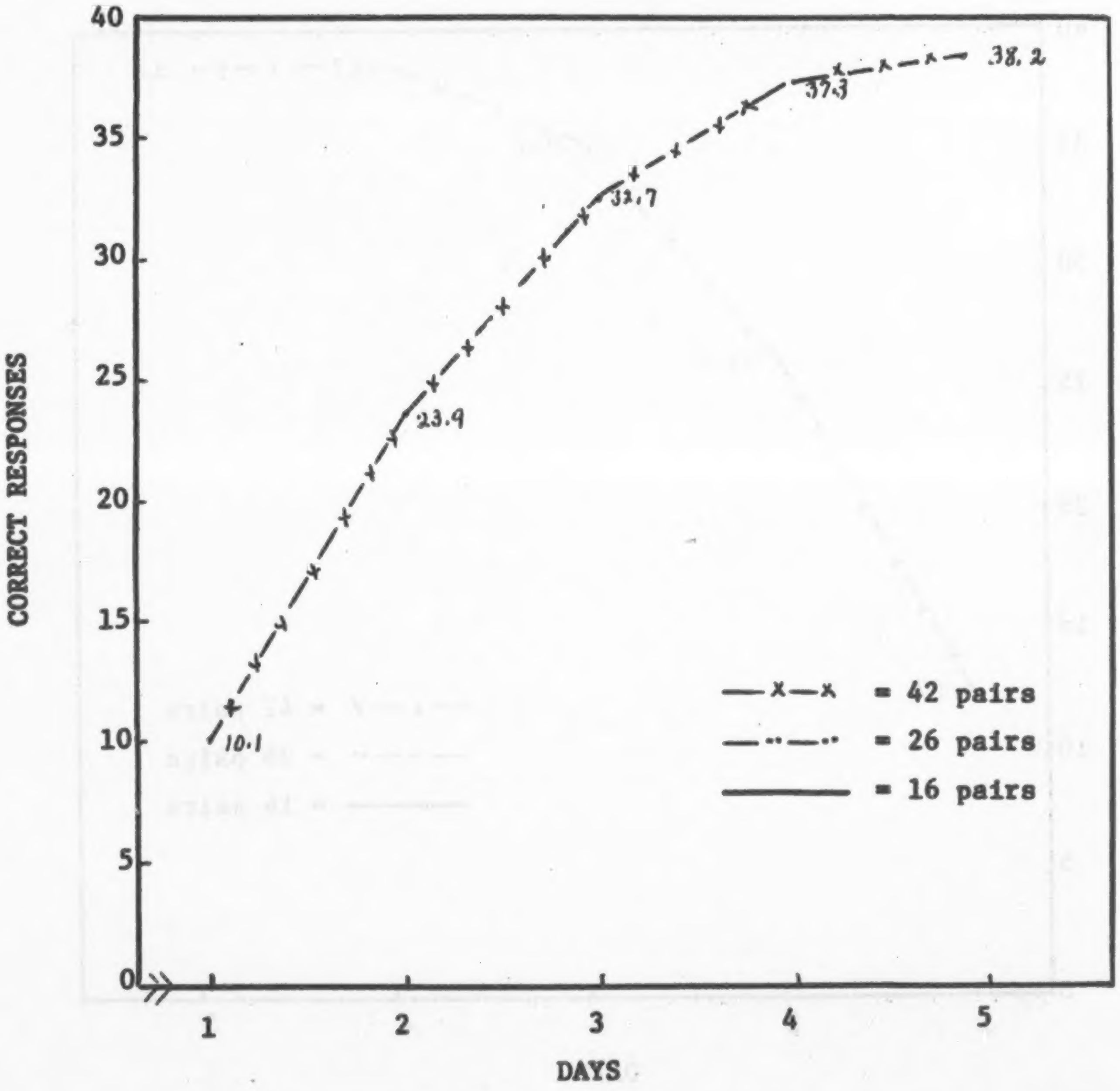
NUMBER OF CORRECT RESPONSES OVER 5 TRIALS ON PRACTICE LIST,

DAYS ONE THROUGH FIVE : GROUP 42 - 16



NUMBER OF CORRECT RESPONSES OVER 5 TRIALS ON PRACTICE LIST,

DAYS ONE THROUGH FIVE : GROUP 42 - 26



Appendix H

Correlations Among All Log_e Learning and Relearning Times: All groups

Correlations Among All Log_e Learning and Relearning Times: Group 16-16

	List 1 Day				List 2 Day			List 3 Day		List 4 Day
	1	2	3	4	2	3	4	3	4	4
List 1 Day 1	1.00	.66	.82	.08	.75	.53	.86	.76	.72	.86
2	0.66	1.00	.74	.39	.86	.59	.85	.88	.51	.81
3	0.82	.74	1.00	.06	.61	.77	.86	.76	.65	.96
4	0.08	.39	.06	1.00	.31	.11	.25	.37	.37	.02
List 2 Day 2	0.75	.86	.61	.31	1.00	.42	.80	.83	.55	.74
3	0.53	.63	.77	.11	.42	1.00	.67	.63	.70	.67
4	0.86	.85	.86	.25	.80	.67	1.00	.85	.81	.91
List 3 Day 3	0.76	.88	.76	.37	.83	.63	.85	1.00	.60	.75
4	0.72	.51	.65	.37	.55	.70	.81	.60	1.00	.66
List 4 Day 4	0.86	.81	.96	.02	.74	.67	.91	.75	.66	1.00

Correlations Among All Log$_e$ Learning and Relearning Times: Group 26-26

	List 1 Day				List 2 Day			List 3 Day		List 4 Day
	1	2	3	4	2	3	34	3	4	4
List 1 Day 1	1.00	.71	.36	.40	.84	.52	.51	.88	-.03	.76
2	.71	1.00	.47	.63	.70	.64	.67	.75	.44	.68
3	.36	.49	1.00	.77	.46	.11	.35	.41	.04	.47
4	.40	.63	.77	1.00	.36	.11	.45	.46	.45	.44
List 2 Day 2	.84	.70	.46	.36	1.00	.80	.40	.92	.02	.96
3	.52	.64	.11	.11	.80	1.00	.34	.78	.35	.80
4	.51	.67	.35	.45	.40	.34	1.00	.65	.17	.48
List 3 Day 3	.88	.75	.41	.46	.92	.78	.65	1.00	.17	.93
4	-.03	.44	.04	.45	.02	.35	.17	.17	1.00	.12
List 4 Day 4	.76	.68	.47	.44	.96	.80	.80	.93	.12	1.00

Correlations Among All Log_e Learning and Relearning Times: Group 42-42

		List 1 Day				List 2 Day			List 3 Day		List 4 Day
		1	2	3	4	2	3	4	3	4	4
List 1	Day 1	1.00	.09	.29	-.40	.69	.12	-.01	.85	.19	.92
	2	.09	1.00	.60	.37	-.03	.28	.42	.14	.65	.36
	3	.29	.60	1.00	-.01	.32	.36	.52	.40	.70	.41
	4	-.40	.37	-.01	1.00	-.24	-.04	.08	-.12	.13	-.25
List 2	Day 2	.69	-.03	.32	-.24	1.00	.17	.07	.89	.26	.74
	3	.12	.28	.36	.04	.17	1.00	.85	.33	.74	.13
	4	-.01	.42	.52	.08	.07	.85	1.00	.16	.63	.00
List 3	Day 3	.85	.14	.40	-.12	.89	.33	.66	1.00	.40	.87
	4	.19	.65	.70	.13	.26	.74	.63	.40	1.00	.37
List 4	Day 4	.92	.36	.41	-.25	.74	.13	.00	.87	.37	1.00

Correlations Among All Log_e Learning and Relarning Times: Group 26-16

		List 1 Day				List 2 Day			List 3 Day		List 4 Day
		1	2	3	4	2	3	4	3	4	4
List 1	Day 1	1.00	.47	.36	-.27	.70	.29	.12	-.03	-.11	.24
	2	.47	1.00	.44	-.09	.59	.13	.06	.57	.42	.71
	3	.36	.44	1.00	-.51	.83	.17	-.53	-.03	.41	.67
	4	-.27	-.09	-.51	1.00	-.43	-.09	.48	-.00	-.33	-.40
List 2	Day 2	.70	.59	.83	0.433	1.00	.27	-.19	-.09	.48	.63
	3	.29	.13	.17	-.09	.27	1.00	.27	.21	.31	.25
	4	.12	.06	-.53	.48	-.19	.27	1.00	.39	-.07	-.03
List 3	Day 3	-.03	.57	-.03	-.00	-.09	.21	.39	1.00	.14	.58
	4	-.11	.42	.41	-.33	.48	.13	-.07	.14	1.00	.58
List 4	Day 4	.24	.71	.67	-.40	.63	.25	-.03	.58	.58	1.00

Correlations Among All $\log_e$ Learning and Relarning Times: Group 42-16

		List 1 Day				List 2 Day			List 3 Day		List 4 Day
		1	2	3	4	2	3	4	3	4	4
List 1	Day 1	1.00	.03	.30	.28	.71	.51	.56	.66	-.26	.12
	2	.03	1.00	-.09	-.02	.46	.43	-.01	.23	-.10	.66
	3	.30	-.09	1.00	.99	.42	.16	.40	.13	.28	.21
	4	.28	-.02	.99	1.00	.39	.15	.39	.14	.22	.22
List 2	Day 2	.71	.46	.42	.39	1.00	.38	.24	.53	-.03	.55
	3	.51	.43	.16	.15	.38	1.00	.72	.60	.04	.26
	4	.56	-.01	.40	.39	.24	.72	1.00	.38	.35	.11
List 3	Day 3	.66	.23	.13	.14	.53	.60	.38	1.00	-.09	.18
	4	-.26	-.10	.28	.22	-.03	.04	.35	-.09	1.00	.31
List 4	Day 4	.12	.66	.21	.22	.55	.26	.11	.18	.31	1.00

Correlations Among All Log$_e$ Learning and Relearning Times: Group 42-26

		List 1 Day				List 2 Day			List 3 Day		List 4 SDay
		1	2	3	4	2	3	4	3	4	4
List 1	Day 1	1.00	.34	.53	.27	.78	.73	.36	.46	.44	.63
	2	.34	1.00	.71	.37	.59	.54	.19	-.14	.63	.27
	3	.53	.71	1.00	.17	.57	.67	.34	.15	.62	.58
	4	.27	.37	.17	1.00	.33	.01	-.69	.16	.27	.17
List 2	Day 2	.78	.59	.57	.33	1.00	.88	.27	.49	.70	.77
	3	.73	.54	.67	.01	.88	1.00	.59	.59	.60	.81
	4	.36	.19	.34	-.69	.27	.59	1.00	.10	.04	.20
List 3	Day 3	.56	-.14	.15	.16	.49	.59	.10	1.00	.12	.79
	4	.44	.63	.62	.27	.70	.60	.04	.12	1.00	.62
List 4	Day 4	.63	.27	.58	.17	.77	.81	.20	.79	.62	1.00

Appendix I

Part V: Modern Language Aptitude Test

(Carroll and Sapon)

Adaptation by David Horton

VOCABULARY LEARNING TEST

(Please Print)

NAME: ______________________ (Last) ______________________ (First) AGE: ____ DATE: __________

SEX: __________ SCORE: __________ CONDITIONS: __________

INSTRUCTIONS

In this test you are to learn the English equivalent of a number of Kurdish words, the following is a sample item.

hib

A. fish

B. stop

C. warm

[✓ D.] ride

E. paper

Since hib means ride you would draw a circle around the letter D.

These questions are to be done from memory.

DO NOT TURN PAGE UNTIL TOLD TO DO SO !!

Instructions. Your task is to MEMORIZE the Kurdish-English vocabulary below. Wait for the signal, then you will be given 1 1/2 minutes to study the vocabulary printed below. At the end of the 1 1/2 minutes the examiner will give you the signal to start filling in the blanks in the PRACTICE EXERCISE SHEET. You are allowed to look back at the vocabulary on this page when you are filling in the blanks on the practice exercise sheet. After filling in the blanks, continue studying if there is still time.

Vocabulary (Memorize for 1 1/2 minutes)

Kurdish	English	Kurdish	English
hib	-ride	birs	-mule
daq	-then	zumap	-many
besus	-noise	blim	-eagle
rintu	-girl	doi	-stop
mu	-night	tyeh	-hard
ugop	-dull	mub	-off
lyof	-stay	poy	-lion
chevad	-try	tocro	-paper
balap	-fear	di	-burn
nerm	-fish	etic	-sky
steva	-arm	i	-cup
noa	-law	hul	-warm

Do not fill in the blanks below until you are told to do so !!
Continue studying.

When you are told to do so, write the English meanings in the spaces following the Kurdish words.

Kurdish	English	Kurdish	English	Kurdish	English
chevad	________	hib	________	birs	________
doi	________	zumap	________	blim	________
ugop	________	mu	________	daq	________
hul	________	nerm	________	etic	________
besus	________	tocro	________	steva	________
noa	________	i	________	poy	________
lyof	________	balap	________	rintu	________
di	________	mub	________	tyeh	________

DO NOT TURN PAGE UNTIL TOLD TO DO SO! CONTINUE STUDYING!!

Draw a circle around the letter of the English word that you select. You are not permitted to look back at the previous page.

1. mub

A. in
B. off
C. then
D. stay
E. law

2 1

A. ball
B. at
C. arm
D. warm
E. cup

3. poy

A. lion
B. mule
C. dull
D. last
E. off

4. lyof

A. try
B. fear
C. stay
D. burn
E. ride

5. etic

A. fish
B. arm
C. eagle
D. sky
E. paper

6. tyeh

A. night
B. hard
C. noise
D. burn
E. cup

7. steva

A. law
B, arm
C. warm
D. noise
E. dull

8. rintu

A. stay
B. fear
C. girl
D. many
E. stop

9. balap

A. map
B. ride
C. warm
D. lion
E. fear

10. blim

A. hard
B. young
C. many
D. burn
E. eagle

11. zumap

A. warm
B. mule
C. cup
D. many
E. fish

12. hul

A. off
B. warm
C. in
D. stop
E. then

13. daq

A. not
B. night
C. then
D. law
E. try

14. mu

A. yes
B. night
C. off
D. sky
E. no

15. noa

A. law
B. ride
C.. try
D. run
E. mule

16. ugop

A. stay
B. girl
C. then
D. dull
E. off

17. tocro

A. never
B. paper
C. fear
D. many
E. burn

18. nerm

A. try
B. eagle
C. night
D. then
E. fish

19. chevad

A. fish
B. lion
C. arm
D. try
E. warm

20. hib

A. stop
B. night
C. hard
D. ride
E. sky

21. birs

A. kite
B. lion
C. mule
D. then
E. burn

22. besus

A. hard
B. warm
C. noise
D. dull
E. paper

23. doi

A. try
B. lion
C. many
D. night
E. stop

24. di

A. burn
B. ride
C. night
D. cup
E. stay

STOP. WAIT FOR FURTHER INSTRUCTIONS.

References

Archer, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. Psychol. Monogr., 1960, 74 (Whole No. 497).

Asch, S. E., and Ebenholtz, S. M. The process of free recall: evidence for non-associative factors in acquisition and retention. J. Psychol., 1962, 54, 3-31.

Asher, J. J. The paired associate task: adequate or inadequate in psycholinguistic research? Paper read at Western Psychological Association, San Francisco, April, 1962.

Barnes, J. M., and Underwood, B. J. Fate of first-list associations in transfer theory. J. exp. Psychol.,1959, 58, 97-105.

Bilodeau, I. M., and Schlosberg, H. Similarity in stimulation conditions as a variable in retroactive inhibition. J. exp. Psychol., 1951, 41, 199-204.

Briggs, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. J. exp. Psychol., 1954, 47, 285-293.

Briggs, G. E., and Thompson, R. F., and Brogden, W. J. Retention functions in reproductive inhibition. J. exp. Psychol., 1954, 48, 419-423.

Briggs, G. E. Retroactive inhibition as a function of degree of original and interpolated learning. J. exp. Psychol., 1957, 53, 60-67.

Brown, W. Whole and part methods in learning. J. educ. Psychol., 1924, 15, 229-233.

Bugelski, B. R., and Cadwallader, T. C. A reappraisal of the transfer and retroaction surface. J. exp. Psychol., 1956, 52, 360-366.

Burke, M. L. Relationships among learning, practice, and recall. Appendix A to Proposal S-332-64 of Project S-169, U. S. Office of Education, 1964.

Carroll, J. B. Programmed self-instruction in Mandarin Chinese. Wellesley, Massachusetts: Language Testing Fund, 1963.

Carroll, J. B. Research on teaching foreign languages. in Gage, N. L. Handbook of research on teaching. New York: Rand McNally, 1963. (A.E.R.A.).

Carroll, J. B., and Burke, M. L. Parameters of paired-associate verbal learning: length of list, meaningfulness, rate of presentation and ability. J. exp. Psychol., 1965, 69, 543-553.

Carroll, J. B., and Sapon, S. M. Modern language aptitude test. New York: The Psychological Corporation, 1958.

Cieutat, V. J., Stockwell, F. E., and Noble, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness (m, m') in paired-associate learning. J. exp. Psychol., 1958, 56, 193-202.

Cofer, C. N. (ed.). Verbal learning and verbal behavior. New York: McGraw-Hill, 1961.

Cook, T. W. Factors in whole and part learning as visually perceived maze. J. genet. Psychol., 1936, 49, 3-31.

Cook, T. W. Whole versus part learning in the spider maze. J. exp. Psychol., 1937, 20, 477-494.

Cooley, W. W., and Lohnes, P. R. Multivariate procedures for the behavioral sciences. New York: Wiley, 1962.

Davis, A. J., and Meenes, M. Factors determining the relative efficacy of the whole and part methods of learning. *J. exp. Psychol*., 1932, 15, 716-727.

Dempster, A. P. Personal communications. Harvard University, Statistics Department, March and April, 1965.

Ebbinghaus, H. Memory: *a contribution to experimental psychology* (1885). transl. Ruger, H. A., and Bussenius, C. E. Columbia Univ. *Coll*. *Educ*. *Reprints*, No. 3. New York: Teachers College, Columbia University, 1913.

Edwards, A. L. Experimental design in psychological research. New York: Holt, Rinehart and Winston, 1962.

Ephrussi, P. Experimentelle Beiträge zur lehre vom Gedachtnis. *Zsch*. *f*. *Psychol*., 1904, *37*, 56-103, 161-182. *in* McGeoch, C. O. Whole-part problem. *Psychol*. *Bull*.,1931, *28*, 713-739.

Fisher, A. R., and Yates, F. *Statistical Tables for biological, agricultural, and medical research*. New York: Hafner, 1953.

French, J. W., Ekstrom, R., and Price, L. A. Test of associative fluency. *in Kit of reference tests for cognitive factors*. Princeton, New Jersey: Educational Testing Service, 1963.

Gillette, A. L. Learning and retention: a comparison of three experimental procedures. *Arch*. *Psychol*., 1936, *28*, No. 198.

Gladis, M., and Braun, H. W. Age differences in transfer and retroaction as a function of intertask response similarity. *J*. *exp*. *Psychol*., 1958, *55*, 25-30.

Greenspoon, J., and Ranyard, R. Stimulus conditions and retroactive inhibition. *J*. *exp*. *Psychol*., 1957, *53*, 55-59.

Higa, M. The concepts of "difficulty" as it relates to the organization of learning material. Unpubl. paper, Graduate School of Education, Harvard Univeristy, 1961.

Higa, M. Interference effects of intralist word relationships in verbal learning. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 170-175.

Hovland, C. I. Human learning and retention. in Stevens, S. S. (ed.). Handbook of experimental psychology. New York: Wiley, 1951.

Hsaio, H. H. An experimental comparison of the whole and part method of learning. in Stevens, S. S. (ed.). Handbook of experimental psychology. New York: Wiley, 1951.

Irion, A. L. The relation of 'set' to retention. Psychol. Rev., 1948, 55, 336-341.

Jenkins, W. O., and Postman, L. An experimental analysis of set in rote learning: retroactive inhibition as a function of changing set. J. exp. Psychol., 1949, 39, 69-72.

Jones, K. Multivariate statistical analyzer. Harvard University, 1964.

Jost, A. in Hovland, C. I. Human learning and retention. in Stevens, S. S. (ed.). Handbook of experimental psychology. New York: Wiley, 1951.

Kjeldergaard, P. M. Predicting paired-associate learning speed. Psychol. Rev., 1962, 11, 353-354.

Krueger, W. C. F. The effect of overlearning on retention. J. exp. Psychol., 1929, 12, 71-78.

Lakenan, M. E. The whole and part method of memorizing poetry and prose. J. educ. Psychol., 1913, 4, 189-198.

Luh, C. W. The conditions of retention. Psychol. Monogs., 1922, 31, No. 3.

McGeoch, J. A. The influence of degree of learning upon retroactive inhibition. Amer. J. Psychol., 1929, 41, 252-262.

McGeoch, G. O. The intelligence quotient as a factor in the whole-part problem. J. exp. Psychol., 1931, 14, 333-358.

McGeoch, G. O. Whole-part problem. Psychol. Bull.,1931, 28, 713-739.

McGeoch, J. A. The psychology of human learning: an introduction. New York: Longmans, Green, 1942.

McGeoch, J. A., and Irion, A. L. The psychology of human learning. New York: Longmans, Green, 1952 (2nd edition).

Melton, A. W. Comments on professor Postman's paper. in Cofer, C. N. (ed.). Verbal learning and verbal behavior. New York: McGraw-Hill, 1961.

Melton, A. W., and Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. Amer. J. Psychol., 1940, 53, 173-203.

Morgan, R. L., and Underwood, B. J. Proactive inhibition as a function of response similarity. J. exp. Psychol., 1950, 40, 592-603.

Nagge, J. W. An experimental test of the theory of associative interference. J. exp. Psychol., 1935, 18, 663-682.

Newmann, G. Experimentelle Beitrage zur Lehre von der Okonomie und technik des lernens. in Lay und Meumann's Die Experimentelle Padagogie, 1977, 4, 63-101, 155-174. in McGeoch, G. O. Whole-part problem. Psychol. Bull.,1931, 28, 713-739.

Newton, J. M., and Wickens, D. D. Retroactive inhibition as a function of the temporal position of the interpolated learning. J. exp. Psychol., 1956, 51, 149-154.

Noble, C. E., and McNeely, D. A. The role of meaningfulness (m) in paired-associate verbal learning. J. exp. Psychol., 1957, 53, 16-22.

Noble, C. E. Meaningfulness (m) and transfer phenomena in serial verbal learning. J. Psychol., 1961 a, 52, 201-210.

Noble, C. E. Measurements of association value (a), rated associations (a'), and scaled meaningfulness (m') for the 2100 CVC combinations of the English alphabet.. Psychol. Rep., Monogs. Suppl. 3-V8, 1961 b.

Orbison, W. D. The relative efficiency of whole and part methods of learning paired-associates as a function of the length of list. Ph.D. thesis (unpublished), Yale University, 1944.

Osgood, C. E. The similarity paradox in human learning: a resolution. Psychol. Rev., 1949, 56, 132-143.

Pechstein, L. A. Whole versus part methods in learning nonsensical syllables. J. educ. Psychol., 1918, 9, 381-387.

Pechstein, L. A. Massed vs. distributed effort in learning. J. educ. Psychol., 1921, 12, 92-97.

Pechstein, L. A. The whole versus part methods in learning: comparison and summary. Stud. in Educ. (Yrbk. Natl. Soc. Coll. Teach. Educ.), 1926, 15, 181-186. in McGeoch, G. O. Whole-part problem. Psychol. Bull., 1931, 28, 713-739.

Postman, L. The present status of interference theory. in Cofer, C. N. (ed.). Verbal learning and verbal behavior. New York: McGraw-Hill, 1961.

Radossawljewitsch, P. R. in Hovland, C. I. Human learning and retention. in Stevens, S. S. (ed.). Handbook of experimental psychology. New York: Wiley, 1951.

Russell, W. A., and Jenkins, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff word association test. Tech. Rep., No. 11, ONR Contract N8 onr-66216, University of Minnesota, 1954.

Sand, M. C. The effect of length of list upon retroactive inhibition when degree of learning is controlled. Arch. Psychol., 1939, 33, No. 238.

Seagoe, M. V. The influence of degree of wholeness on whole-part learning. J. exp. Psychol., 1936, 19, 763-768.

Seagoe, M. V. Qualitiative wholes: a re-evaluation of the whole-part problem. J. educ. Psychol., 1936, 27, 537-545.

Seibert, L. C. A series of experiments on the learning of French vocabulary. Johns Hopkins University Studies in Education, No. 18, 1932.

Suppes, P. Problems of optimization in learning a list of simple items. Psychology Series, Stanford University, (Technical Report No. 57), 1963.

Suppes, P. Modern learning theory and the elementary-school curriculum. Amer. Educ. Res. I., 1964, 1, 79-94.

Thorndike, E. L., and Lorge, I. The teachers' word book of 30,000 words. New York: Teachers College, Columbia University, 1944.

Thorndike, R. L. Test of verbal intelligence. in Minor, J. B. Intelligence in the United States. New York: Springer, 1957, 53.

Thune, L. E., and Underwood, B. J. Retroactive inhibition as a function of degree of interpolated learning. *J*. *exp*. *Psychol*., 1943, *32*. 185-200.

Thune, L. E. The effect of different types of preliminary activities on subsequent learning of paired-associate material. *J*. *exp*. *Psychol*., 1950, *40*, 423-438.

Twedt, H. M., and Underwood, B. J. Mixed vrs. unmixed lists in transfer studies. *J*. *exp*. *Psychol*., 1959, *58*, 111-116.

Twining, P. E. The relative importance of intervening activity and lapse of time in the production of forgetting. *J*. *exp*. *Psychol*., 1940, *26*, 483-501.

Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. 1945, *Psychol*. *Monogr*., 59, No. 3.

Underwood, B. J. "Spontaneous" recovery of verbal associations. *J*. *exp*. *Psychol*., 1948, *38*, 429-439.

Underwood, B. J. *Experimental* *psychology*. New York: Appleton-Century-Croft, 1949.

Underwood, B. J. Proactive inhibition as a function of time and degree of prior learning. *J*. *exp*. *Psychol*., 1949, *39*, 24-34.

Underwood, B. J. Studies of distributed practice: XII. Retention following varying degrees of interpolated learning. *J*. *exp*. *Psychol*., 1954, *47*, 294-300.

Underwood, B. J. An evaluation of the Gibson theory of verbal learning. *in* Cofer, C. N. (ed.). *Verbal* *learning* *and* *verbal* *behavior*. New York: McGraw-Hill, 1961.

Underwood, B. J., and Richardson, J. Studies of distributed practice: XVII. Interlist interference and the retention of paired consonant syllables. J. exp. Psychol., 1957, 54, 274-279.

Underwood, B. J., and Schulz, R. W. Meaningfulness and verbal learning. Chicago: Lippincott, 1960.

Underwood, B. J., and Schulz, R. W. Studies of distributed practice: XX. Sources of interference associated with differences in learning and retention. J. exp. Psychol., 1961, 61, 228-235.

Youtz, A. C. An experimental evaluation of Jost's laws. Psychol. Monogr., 1941, 53, No. 1.

Waters, R. H. Degree of learning and proactive inhibition in retention. Psychol. Bull., 1942, 39, 495-496.

Woodworth, R. S., and Schlosberg, H. Experimental psychology (revised edition). New York: Henry Holt, 1955.

VITA

Mary Long Burke

1956-1958	Our Lady of Cincinnati College Cincinnati, Ohio	
1958-1960	Trinity College Washington, D. C. Major in English and Creative Writing	A.B.
1960-1961	Radcliffe College Major in English	A.M.T.
1961 (second semester)	Wellesley Senior High School Intern in English	
1961-present	Harvard Graduate School of Education	Candidate for Ed. D.
1961-present	Harvard Graduate School of Education Research assistant to Dr. John B. Carroll 1/2 time to 1964 3/5 time, under U.S. Office of Education grant, 1964-1965	
1962 summer	Nazareth College Louisville, Kentucky Lecturer in English Taught American Literature and Advanced Creative Writing	
1963 summer	The RAND Corporation Santa Monica, California Consultant (in Computer Simulation of Verbal Behavior)	